

**FINAL REPORT**

# **Quantifying Key Drivers of Climate Variability and Change for Puerto Rico and the Caribbean**

**Katharine Hayhoe, Texas Tech University (PI)**

With contributions from Jung-Hee Ryu, Anne Stoner, and the TTU High Performance Computing Center

---

Agreement Number G10AC00582

Date of Report: April 28, 2013

Period of Time Covered by Report: October 1, 2011 to September 30, 2012

Actual total cost: \$149,368.00



## **PUBLIC SUMMARY**

One of the most important reasons we study climate change is to understand how it may exacerbate existing vulnerabilities in both human systems and the natural environment. We know that coastal flooding and storm damage will occur as a result of rising sea level and increasingly more powerful tropical cyclones and hurricanes. In the case of Puerto Rico and other Caribbean islands, however, other key vulnerabilities to climate change relate to the issue of how to sustain growing populations and unique ecosystems with limited land area and scarce water resources.

This project laid the foundation to explore the potential impacts of climate change on Puerto Rico and other Caribbean islands. First, we quantified the ability of global climate models to reproduce the large-scale atmosphere and ocean dynamics that control temperature and rainfall variability in the Caribbean. Then, we generated high-resolution projections of daily maximum and minimum temperature and 24h cumulative precipitation for over 200 long-term weather stations. These simulations capture the range of changes projected by 32 different climate models under a higher and lower emissions future. Next, for Puerto Rico, we derived and analyzed projected changes in 85 secondary climate indicators, including seasonal averages, thresholds, and extremes.

Puerto Rico is expected to warm faster than the global average, with increases in both mean and extreme temperatures, including days per year over 95°F and nights warmer than 85°F. Rainfall is projected to decrease, particularly in the wet season, with more frequent dry days. The frequency of “moderate extreme” precipitation (e.g. more than 1 inch of rain) is projected to decrease, while more extreme precipitation (e.g. more than 3 inches of rain in a day) is expected to become more common. Projected temperature changes are large enough to affect temperature-sensitive crops, species, and ecosystems, while the combined effects of changes in temperature and precipitation are likely to increase the demand for energy, the risk of water stress and drought, and the risk of impacts from heavy rainfall events.

## TABLE OF CONTENTS

TECHNICAL SUMMARY .....	5
PURPOSE AND OBJECTIVES .....	8
ORGANIZATION AND APPROACH .....	9
Historical and Future Climate Scenarios.....	10
Global Climate Model Simulations .....	12
Downscaling Model .....	15
Gridded Observations .....	19
Station Observations .....	20
Computational Methods .....	21
PROJECT RESULTS .....	21
Step One: High-Resolution Projections for the Caribbean.....	21
Step Two: Evaluation of Global Model Performance over the Caribbean.....	25
Step Three: Analysis of Climate Change Indicators for Puerto Rico .....	26
ANALYSIS AND FINDINGS .....	28
Global Model Performance: Caribbean Temperature.....	28
Global Model Performance: Caribbean Precipitation .....	32
Climate Change Projections for Puerto Rico.....	35
CONCLUSIONS AND RECOMMENDATIONS.....	44
OUTREACH .....	46
REFERENCES .....	47
APPENDIX A: Latitude, Longitude and Station ID for High-Resolution Climate Projections	
APPENDIX B: Analysis of GCM ability to simulate Caribbean precipitation drivers <i>“Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations,”</i> Ryu & Hayhoe 2013	
APPENDIX C: Climate Projections for Puerto Rico: Secondary Indicator Plots	
APPENDIX D: Description and Evaluation of the ARRM Downscaling Method <i>“An asynchronous regional regression model for statistical downscaling of daily climate variables,”</i> Stoner et al. 2012	

## GUIDELINES FOR USE OF CARIBBEAN HIGH-RESOLUTION CLIMATE PROJECTIONS

This project has conducted a comprehensive dynamical analysis of the ability of 32 different global climate models to simulate observed temperature and rainfall variability over the Caribbean, used these models to generate future projections of temperature and precipitation, and calculated projected changes in 85 secondary climate indicators for long-term weather stations in Puerto Rico.

Based on this analysis, we conclude the following:

- For **temperature**, there is no set of generally “better-performing” models. All models display a lag in seasonal air and sea surface temperature that is seen in both higher- and lower-resolution models as well as in both atmosphere-only and coupled atmosphere-ocean models. *The recommended course of action is to use as many global climate models as is feasible for any given impact analysis.*
- For **precipitation**, projected changes in seasonal precipitation and precipitation extremes simulated by models that are able to reproduce the observed large-scale dynamics that control precipitation over the Caribbean are noticeably different than projected changes from models that are not. *The recommended course of action is to use global climate models that have been proven capable of simulating the processes that drive changes in precipitation over the Caribbean.* A list of these models is provided in Table 3.
- For **precipitation**, there is a larger difference between seasonal and extreme precipitation projections simulated by “good” CMIP3 vs. CMIP5 models than between the projections simulated by the “good” models vs. “poor” models. It is likely (although not yet proven) that these differences are due to CMIP3-based projections primarily being driven by increases in carbon dioxide, while CMIP5-based projections are driven by changes in both carbon dioxide and aerosols, which can affect cloud properties and hence rainfall. Because of the substantial differences in the projected changes in wet season and extreme rainfall between CMIP3 vs. CMIP5 simulations, *the recommended course of action is to preferentially use CMIP5 simulations only for any projections of changes in wet season rainfall or extreme rainfall.*

In addition to these Caribbean-specific recommendations, we also recommend the following best practices:

- Do not attempt to select a single “best” model by comparing biases in seasonal temperature or precipitation for a location of interest. This evaluation method offers no guarantee that the top-performing models in terms of historical biases will also be the top-performing models in terms of simulating the impacts of global change on the same location.
- When using multiple climate model simulations as input to an impact analysis, always average across climate models as the very last step in the analysis. Unless the relationships between climate change and the impacts being studied are entirely linear, averaging across climate models too early in the analysis will artificially average across, and remove, the temporal variability from the climate projections, leading to incorrect results.
- Do not average across multiple future scenarios. Scenarios are not like physical systems, where averages may yield the most likely value. Scenarios are entirely separate, independent pictures of what the future may look like, given a set of assumptions regarding socio-economic and technological development. Results of any analysis can be averaged across climate models, but should be presented independently for each scenario.
- There is no one most likely future scenario. It is impossible to predict human behavior. The most frequently recommended course of action is to consider the impacts resulting from a higher vs. a lower future scenario, as this will cover the range of projected changes from the full range of scenarios.
- Climate simulations are intended to match observations over climate time scales of decades, not days. Do not expect a climate simulation to match day-to-day observations at any given location. The averages should match over 20-30 years, but climate models are allowed to develop their own unique patterns of day-to-day climate variability.
- Uncertainty in future projections is the result of multiple factors. Over the next decade or two, the most important source of uncertainty in future projections is natural variability. Scientific uncertainty in the response of the climate system to human activities is the main source of uncertainty in temperature over the next few decades, and in precipitation through the end of the century. Human or scenario uncertainty (what our emissions will be) is the largest source of uncertainty in temperature past mid-century.

## TECHNICAL SUMMARY

This project laid the foundation of data and analyses required to assess the potential impacts of changes in temperature and precipitation for Puerto Rico and other Caribbean islands. The first step in this project was to compile, standardize, and (for station data) quality-control available observations for the region. These consisted of:

- Station-based observations of daily maximum and minimum temperature (27 locations in Puerto Rico and a total of 284 for maximum and 282 for minimum temperature in the Caribbean) and 24h cumulative precipitation (77 locations in Puerto Rico and a total of 231 in the Caribbean) from the Global Historical Climatology Network, National Climatic Data Center, and U.K. Met Office Land Surface Station datasets (variable record length)
- Assimilated (station observations + satellite) monthly mean precipitation covering the entire Caribbean, gridded at a resolution of 2.5 degrees, from the Global Precipitation Climatology Project (1979-2008)
- Reanalysis-based 2-meter monthly mean temperature and upper air fields (geopotential height, winds, vertical velocity and specific humidity) covering the entire Caribbean, gridded at a resolution of 1.4 degrees, from the European Centre for Medium-Range Weather Forecasts (1957-2002)
- Monthly mean sea surface temperature covering the entire Caribbean, gridded at a resolution of 1 degree, from the National Oceanic and Atmospheric Administration (1982-2008)
- Ocean mixed layer depth, gridded at a resolution of 0.25 degrees, from the Levitus World Ocean Atlas (1900-1992)

The next step was to obtain and process global climate model (GCM) simulations of relevant variables for the Caribbean. Two sets of GCM simulations were used, as follows:

- Coupled Model Intercomparison Project version 3 (CMIP3) simulations of daily maximum and minimum temperature and precipitation and monthly sea surface temperature, specific humidity, winds, and geopotential height from 16 global climate models for historical and a range of future emission scenarios, from the World Climate Research Program (1960-2099)
- Coupled Model Intercomparison Project version 5 (CMIP5) simulations of the same variables from 16 global climate models for historical and a range of future concentration scenarios, from the World Climate Research Program (1950-2100)

Since precipitation over the Caribbean is controlled by a complex pattern of large-scale atmosphere and ocean dynamics, it was important to identify the main features affecting variability in temperature and precipitation over the region (using the observed data sources listed above) and assess the degree to which global climate models were able to reproduce these features.

Temperature varies smoothly over the course of the year and is closely related to sea surface temperature. Nearly every global model shows a significant lag in the seasonal cycle of both sea surface temperature and air temperature, ranging from a week to a month compared to observations. Our initial hypothesis that this lag was related to biases in

model simulation of the ocean mixed layer was not substantiated by analysis of the 7 models that had archived ocean mixed layer depth. Some improvement is seen when comparing higher- to lower-resolution models from the same modeling group, atmosphere-only to fully atmosphere-ocean coupled models, and newer (Coupled Model Intercomparison Project version 5) to older (version 3) models. Overall, however, this lag appears to be a general feature of global models primarily related to limitations in their ability to simulate seasonal shifts in atmospheric, rather than ocean, circulation.

Precipitation in the Puerto Rico and the central Caribbean is characterized by a summer wet season ranging from May to November, punctuated by a mid-summer drought (MSD). In contrast to temperature, we found a clear difference in model ability to simulate the seasonal cycle of precipitation. This difference was directly related to the ability of the models to simulate the eastward extension of a warm pool of sea surface temperature (initiating and ending the wet season) and the mid-summer extension of the semi-permanent North Atlantic Subtropical High pressure system over the region (which temporarily suppresses convection, creating the MSD). Both CMIP3 and CMIP5 GCMs fall into three categories: (1) “good” models able to simulate the timing and magnitude of both of these features; (2) “fair” models with too-strong and too-early shift of NASH over the region, suppressing the early part of the wet season and creating an overly long MSD; and (3) “poor” models with a too-weak NASH, unable to simulate the MSD at all. CMIP3 models were evenly divided between these three groups while the majority of CMIP5 models were in Group 1, suggesting that model ability to simulate these important large-scale drivers of precipitation over the Caribbean has generally improved over time.

We then statistically downscaled outputs from 16 different CMIP3 and 16 more CMIP5 global climate models to each of the long-term station locations in the Caribbean to generate daily simulations of maximum and minimum temperature and 24h cumulative precipitation for the periods 1960 to 2099 (CMIP3) and 1950 to 2100 (CMIP5). The statistical properties of simulated time series are trained to match observed conditions at each station for the historical period, then diverge to capture the characteristics of a higher vs. a lower emissions future over the coming century. For CMIP3, future scenarios consist of the Special Report on Emission Scenarios (SRES) higher A1fi, mid-high A2, mid-low A1B and lower B1 scenarios. For CMIP5, future scenarios consist of the Representative Concentration Pathways (RCP) higher 8.5 and lower 4.5 pathways.

Finally, we calculated a series of 85 secondary climate indicators for Puerto Rico stations, including seasonal changes, thresholds, and extremes. These future projections were analyzed in terms of model performance and changes projected for a range of global mean temperature targets, from +1 to +3°C relative to 1971-2000. Projected changes were divided into two regions for temperature (hot coastal and more temperate inland) and three regions for precipitation (dry northern coast, dry southern coast, and wet inland locations).

With just one degree increase in global temperature, 60% of the wet seasons are projected to be warmer than the historical maximum and, on average, there would be 100 more days over 85°F, 150 more days over 90°F and 35 more days over 95°F each year. With a two-degree increase in global temperature, every day would be warmer than the historical median, 350 days per year will be warmer than the historical 1-in-4 warmest days and 300

days per year will be warmer than the historical 1-in-10 warmest days. For a global mean temperature increase of three degrees, Puerto Rico's average daytime maximum temperature is projected to increase by up to +7°C in the dry season and +6°C in wet season. Increases are projected to be greater for inland locations as compared to coastal, and for nighttime temperatures (over +8°C) compared to daytime. Per degree global mean temperature change, temperature on the warmest day of the year is projected to increase by +3°C while cooling degree-days (a measure of air conditioning demand) are projected to increase by +600. The range of daily temperature is expected to increase, particularly in the wet season.

Rainfall is projected to decrease, particularly in the wet season. More dry days and longer stretches of days without rain are also projected in the wet season. Dry years are expected to become more frequent. Days with "moderate extreme" precipitation (e.g. more than 1 inch of rain) are projected to decrease, but days with more extreme precipitation (e.g. more than 3 inches of rain in a day, or rainfall amounts exceeding the historical 1-in-100 and 1-in-1000 wettest day) are expected to increase.

Projections from the CMIP3 models based on the SRES scenarios and the CMIP5 models based on the RCP scenarios show similar increases in temperature and decreases in precipitation, increasing confidence in the direction and approximate magnitude of future changes for Puerto Rico. For CMIP3, models in Group 1 (good) simulated notably greater decreases in wet season precipitation and increases in dry days and drought periods, as well as smaller changes in daytime maximum temperatures and greater changes in nighttime minimum temperatures, as compared to models in groups 2 or 3. For CMIP5, however, models in Group 1 simulated smaller decreases in wet season precipitation as compared to models in Groups 2 or 3.

Drawing on the work of colleagues at GFDL (Vecchi et al., 2012), it appears that this difference between precipitation decreases simulated by "good" CMIP3 vs. CMIP5 models may be due to the fact that CMIP3-based precipitation changes are primarily driven by carbon dioxide increases, while CMIP5-based precipitation changes are responding to both increases in carbon dioxide *and changes in aerosol emissions and transport, particularly over the Caribbean*. Over this region specifically, the net effect of (a) improved aerosol representation in CMIP5 models, combined with (b) large reductions in aerosol loading amounts in the RCP scenarios appears to somewhat mitigate the drying effects of carbon dioxide on precipitation. If true, the implication is that as aerosol emissions are eliminated, the effects of carbon dioxide may become more evident, leading to further decreases in Caribbean rainfall over longer time horizons than those considered here.

With significant increases in temperature and moderate decreases in precipitation projected for the Caribbean as a whole and Puerto Rico specifically, these climate change projections have important implications: for local agriculture and food supply, temperature-sensitive species and ecosystems, supply and demand of energy (for air conditioning), human health (related to extreme heat and air pollution), availability and demand for water, including increased risk of water stress and drought, and potential infrastructure impacts from increases in the frequency of the most heavy rainfall events. The projections generated by this work are intended to be used to assess the magnitude of future impacts and inform robust adaptation planning.

## PURPOSE AND OBJECTIVES

The first purpose of this work was to create a reliable set of state-of-the-art climate projections that can be used to assess the impacts of climate change on Puerto Rico and the Caribbean. This dataset was created by compiling all available observations for the region (including station observations, satellite assimilation, and reanalysis products), identifying key large-scale drivers of variability in temperature and precipitation across the region, quantifying the ability of global climate models (GCMs) to reproduce these large-scale drivers and their impacts on precipitation and temperature, statistically downscaling simulations from 16 CMIP3 and 16 CMIP5 models to all long-term station records in the Caribbean, calculating projected changes in 85 temperature and precipitation indicators for stations in Puerto Rico, and analyzing results in terms of projected global change and model performance. This dataset enables consistent assessments of climate impacts on both human and natural systems, based on the same common data set. Such analyses produce consistent results that can be compared across the Caribbean and (in conjunction with previous work) across the continental U.S.

The second purpose of this work was to evaluate the hypothesis that global model performance may affect the magnitude and/or direction of projected change. This hypothesis was tested by identifying key features of regional climate that affect air temperature and rainfall, including the Caribbean Low-Level Jet, the North Atlantic Subtropical High, the sea surface temperature warm pool, ocean mixed layer depth, and patterns of moisture convergence and divergence across the region. Global models were then compared with observed patterns of variability and change to determine whether future projections could be differentiated between those models able to simulate observed large-scale circulation features and those that were not. For both CMIP3 and CMIP5 simulations, there was a noticeable difference between the ensemble mean projections from better vs. poorer-performing models for wet season precipitation and, to a lesser degree, wet-season temperature. This suggests that model performance *can* affect the magnitude of future projections and should be taken into account when assessing the impacts of climate change in the May-November Caribbean wet season.

All the objectives of the original work were met. The extent of the original analysis was also **expanded in three important ways**: first, by incorporating the latest projections from CMIP5 models (first made available in 2012-2013) and comparing the results of these models with those obtained from earlier CMIP3 analyses (first made available circa 2005); second, by extending the original proposed evaluation of global climate models, which initially focused on teleconnection patterns, to instead identify and evaluate regional dynamics, a much more technical and detailed analysis; and third, by calculating 85 secondary climate indicators for Puerto Rico and analyzing projected changes in these indicators under global mean temperature increases of +1, 2, and 3°C relative to 1971-2000 for both hot, dry coastal regions and more temperate, wetter inland areas.

## ORGANIZATION AND APPROACH

The research process of this project consists of five steps:

1. **Obtaining, processing, and quality-controlling the observational inputs and global model simulations**
2. **Analyzing the observations to determine the primary large-scale influences on temperature and precipitation in the Caribbean**
3. **Evaluating global climate model simulations to quantify their ability to simulate these large-scale drivers and their influence on regional temperature and precipitation**
4. **Generating high-resolution projections of daily maximum and minimum temperature and 24h cumulative precipitation for long-term weather stations in the region**
5. **Calculating secondary climate indicators and analyzing projected climate change for Puerto Rico.**

**STEP ONE. Compilation of observations and model simulations.** Climate analysis in the Caribbean is complicated by the relative sparseness of long-term data records (in no small part due to the limited land area in the region) as compared to the continental U.S. For that reason, the first and most essential step in this project was to beat the bushes for any and all available data sources for this region. Database and internet searches, literature reviews, and personal recommendations were used to assemble a set of long-term station observations and available gridded fields for the region derived from satellites, ocean buoys, and reanalysis models. For station observations, an additional quality-control process (developed as part of a simultaneous U.S. downscaling effort) was applied to identify outlier and erroneous values that would compromise the quality of the statistical downscaling model. The observations and model simulations used here are described in detail later in this section.

**STEP TWO. Analysis of large-scale climate drivers.** A key hypothesis of this project was whether global climate model performance affected the magnitude and/or direction of projected future change. Before this hypothesis could be evaluated, it was necessary to identify the primary natural drivers of temperature and precipitation variability over the region. This was an iterative and exhaustive process that included detailed literature reviews of previous research and analysis of multiple surface, ocean, and upper-air fields both over the Caribbean as well as remote influences such as ENSO. Many rabbit-trails were followed in order to eliminate less important features and identify a final set of the most relevant large-scale features to temperature and precipitation variability in the Caribbean. This work is described in more detail in the **ANALYSIS AND RESULTS** section.

**STEP THREE. Evaluation of global climate models.** Evaluation of global climate models was complicated by the limited nature of output fields available. Particularly in the CMIP3 archive, not every model had archived the monthly mean upper air or ocean fields required to assess model ability to simulate the features identified in Step 2; hence, this step included much cobbling together of available resources and envisioning multiple ways to ask the same questions in order to maximize available output from each model. To ensure consistency, the global models were evaluated in the same manner as observations, with simulations often normalized to highlight biases in anomalies, or variability over time.

**STEP FOUR. Generating high-resolution climate projections.** Creating a set of high-resolution climate projections requires global climate model simulations and observations at the desired spatial and temporal scale. Here, we used GCM simulations from CMIP3 and CMIP4 and station-based observations from a range of sources (see observations section, below). This step also required a statistical downscaling model to translate the GCM simulations into regionally-relevant information. To this end, we used a model capable of resolving daily extremes, the Asynchronous Regional Regression Model. The historical and future emission scenarios, GCM inputs, observations, and downscaling model are all described in more detail in this section, below.

**STEP FIVE. Analyzing climate change for Puerto Rico.** The final step in the project was to analyze projected changes in daily maximum and minimum temperature and 24h cumulative precipitation for Puerto Rico, and determine whether model performance affected the magnitude of future change. To this end, an indicator code was first written to derive 85 temperature and precipitation-related indicators from the raw daily data, and an analysis code was then written and applied to average projected changes across: (1) 20-year time periods corresponding to global mean temperature increases of +1, 2, and 3°C relative to the 1971-2000 average, (2) across two regions for temperature (hot coastal vs. temperate inland) and three for precipitation (dry northern coast, dry southern coast, and wet central region), and (3) across good, fair and poor-performing GCMs. The resulting multi-model ensemble and temporal averages were then plotted in Excel and saved as images (see **ANALYSIS AND RESULTS** section) to visualize projected changes.

### Historical and Future Climate Scenarios

**This project used SRES emission scenarios A1fi (higher), A2 (mid-high), A1B (mid-low) and B1 (lower) and the RCP concentration pathways 8.5 (higher) and 4.5 (lower). These scenarios were chosen because they cover a broad range of plausible futures in terms of human emissions of carbon dioxide and other radiatively-active species and resulting impacts on climate.**

Climate model simulations begin with a long, multi-century “*control*” run where external forcing conditions including greenhouse gas concentrations, solar radiation, and volcanoes are fixed at constant levels corresponding to a specific year, generally in the 19<sup>th</sup> century. The choice of control year varies from one modeling group to the next, but is typically between 1850 and 1890. This long run is required for the ocean and atmospheric components of the model to equilibrate with each other and reach a stable climate. Output from control runs was not used in this project.

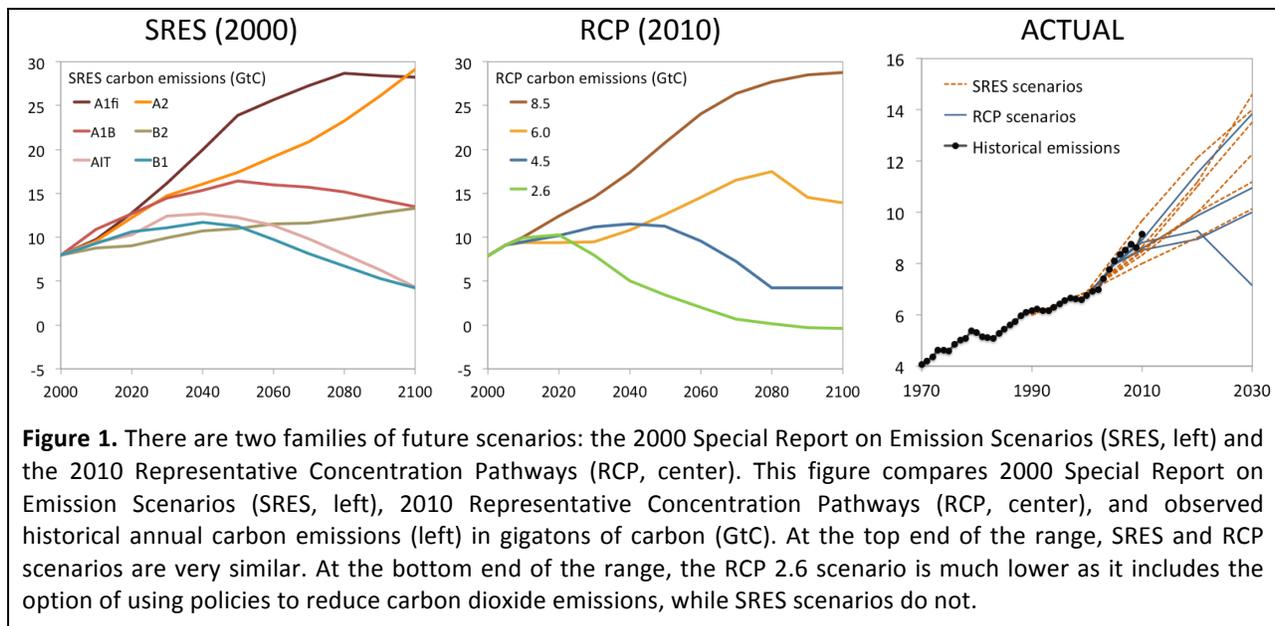
Once climate conditions are stabilized, the output from the control run can be used as input to a *transient historical simulation*. During a transient simulation, the external forcings (including greenhouse gas concentrations, solar radiation, and volcanic eruptions) change from year to year consistent with observed values for that year. The transient historical forcings used by the GCM simulations in this project are the Coupled Model Intercomparison Project’s “20th Century Climate in Coupled Models” or 20C3M total forcing scenarios (Meehl et al. 2007; Taylor et al., 2012). These scenarios include forcing from anthropogenic or human emissions of greenhouse gases, aerosols, and reactive species; changes in solar output; particulate emissions from volcanic eruptions; changes in tropospheric and stratospheric ozone; and other influences required to provide a complete picture of the climate over the last century. As such, these simulations provide the closest

approximation to actual climate forcing from the beginning of the transient experiment to the year 2000. Where multiple 20C3M simulations were available, the first was used here (“run 1”) unless complete daily outputs were not available for that simulation, in which case the next available was used.

In the same way as the control run can provide the starting conditions for multiple historical transient simulations, the historical transient simulation provides the starting conditions for multiple future simulations. To ensure the accuracy of the historical total forcing scenarios, it is customary in the climate modeling community for historical simulations to end at least 5 years before the actual year in which the simulation is being conducted. So although the CMIP3 GCM simulations were typically conducted after 2005, CMIP3 historical total-forcing scenario ends and “future” scenarios begin in 2000. CMIP5 historical scenarios end in 2005 and “future” scenarios begin in 2006. In the future scenarios, most external natural climate drivers are fixed, and human emissions correspond to a range of plausible pathways rather than observed values.

The CMIP3 scenarios used here are those described in the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES; Nakićenović et al., 2000). These scenarios describe internally consistent pathways of future societal development and corresponding emissions. The carbon emissions and global temperature change that result from the SRES scenarios are shown in Figure 1 (left side).

At the higher end of the range, the SRES higher-emissions or fossil fuel intensive scenario (A1FI, for *fossil-intensive*) represents a world with fossil fuel-intensive economic growth and a global population that peaks mid-century and then declines. New and more efficient technologies are introduced toward the end of the century. In this scenario, atmospheric CO<sub>2</sub> concentrations reach 940 parts per million by 2100, more than triple pre-industrial levels of 280 ppm. At the lower end, the SRES lower-emissions scenario (B1) also represents a world with high economic growth and a global population that peaks mid-century and then declines. However, this scenario includes a shift to less fossil fuel-



intensive industries and the introduction of clean and resource-efficient technologies. Emissions of greenhouse gases peak around mid-century and then decline. Atmospheric carbon dioxide levels reach 550 parts per million by 2100, about double pre-industrial levels. Associated temperature changes by end-of-century range from 4 to 9°F based on the best estimate of climate sensitivity.

For this project, climate projections were based on the A1FI higher (dark red), A2 mid-high (orange), A1B mid-low (red) and B1 (blue) lower scenarios. Due to the decision of IPCC Working Group 1 to focus on the A2, A1B and B1 scenarios, only four GCMs had A1FI scenarios available. For other models, daily outputs were not available for all scenarios. Table 1, in the next section on **Global Climate Model Simulations**, summarizes the combinations of GCM simulations and emission scenarios used in this work.

In 2009, the IPCC released a new set of scenarios, called *Representative Concentration Pathways* or RCPs (Moss et al., 2010). In contrast to the SRES scenarios, the RCPs are expressed in terms of carbon dioxide concentrations in the atmosphere, rather than direct emissions. The RCP scenarios are also named in terms of their change in radiative forcing (in watts per meter squared) by end of century: +8.5 W/m<sup>2</sup> and +4.5 W/m<sup>2</sup>.

RCP scenarios can be converted “backwards”, into the range of emissions consistent with a given concentration trajectory, using a carbon cycle model (Figure 1, center). Four RCP scenarios were developed to span a plausible range of future carbon dioxide concentrations, from lower to higher. At the higher end of the range, atmospheric carbon dioxide levels under the RCP 8.5 scenario reaches more than 900 parts per million by 2100. At the lowest, under RCP 2.6 policy actions to reduce carbon dioxide emissions *below zero* before the end of the century (i.e. to the point where humans are responsible for a net uptake of carbon dioxide from the atmosphere) keeps atmospheric carbon dioxide levels below 450 parts per million by 2100. Associated temperature changes by end-of-century range from 2 to 8°F based on the best estimate of climate sensitivity.

For this project, climate projections were based on the RCP 8.5 higher (dark red) and 4.5 lower (blue) scenarios, as these closely match the SRES A1fi and B1 scenarios. Although the CMIP5 archive contains simulations from over 40 models, a much smaller subset (only 16 individual models, from 13 modeling groups) archived daily temperature and precipitation for both the RCP 8.5 and 4.5 scenarios. These models are summarized in Table 1, in the next section on **Global Climate Model Simulations**.

As the SRES scenarios begin in 2000 and the RCP scenarios in 2006, projected carbon emissions can be compared with actual emissions, shown in Figure 1 (right). Currently, actual emissions are near the top of the range of both SRES and RCP scenarios, and are projected to exceed this range by 2030 if current growth rates (averaging more than 3% per year since 2000) continue.

### Global Climate Model Simulations

**To generate high-resolution daily projections for Puerto Rico and the Caribbean, this project used CMIP3 global climate model simulations from 16 different models, and CMIP5 simulations from 16 different models. To analyze GCM ability to simulate observed drivers of monthly temperature and precipitation over the Caribbean, this project used 18 CMIP3 GCMs and 26 CMIP5 GCMs.**

Global climate model simulations, while in a state of constant flux within a given research group or laboratory, are archived at regular intervals by the World Climate Research Programme's Working Group on Coupled Modelling (WGCM). In preparation for the IPCC's Fourth Assessment Report (AR4), the WGCM requested that the US DOE-funded Program for Climate Model Intercomparison and Diagnosis (PCMDI) collect model output from climate modeling centers around the world. This first collection, assembled between 2005 and 2006 and archived by PCMDI, represents models that contributed to phase 3 of the Coupled Model Intercomparison Project (CMIP3; Meehl et al., 2007). These are the results presented in the 2007 IPCC Fourth Assessment Report (AR4).

The CMIP3 GCM simulations used in this project consist of all model outputs archived by PCMDI with daily maximum and minimum temperature and precipitation outputs. Additional simulations were obtained from the archives of the Canadian Centre for Climate Modeling and Analysis, the Geophysical Fluid Dynamics Laboratory, the National Center for Atmospheric Research, and the U.K. Meteorological Office. A total of 17 GCMs met this data-based criteria. The full list of GCMs used, their origin, the scenarios available for each, and the time periods covered by their output are given in Table 1. Output from 12 GCMs was available for the full time period (1960 or 1961 to 2099) while output from 5 more GCMs was available for three time slices (1961-2000, 2046-2065, 2081-2100).

From 2011 through the end of 2012, PCMDI began to collect and archive new GCM simulations that contributed to the fifth phase of CMIP and which will be used in the upcoming IPCC Fifth Assessment Report (AR5; Taylor et al. 2012). The CMIP3 and CMIP5 archives are similar in that most of the same international modeling groups contributed to both. Both provide daily, monthly, and yearly output from climate model simulations driven by a wide range of future scenarios. However, the archives are also different from each other in three key ways. First, many of the CMIP5 models are new versions or updates of previous CMIP3 models and some of the CMIP5 models are entirely new. Some of the CMIP5 models are "Earth System Models" that include both traditional components of the CMIP3 Atmosphere-Ocean General Circulation Models as well as new components such as atmospheric chemistry or dynamic vegetation. Second, the CMIP5 simulations use the RCP scenarios as input for future simulations while the CMIP3 simulations use the SRES scenarios as input (Figure 1). Third, the CMIP5 archive contains many more output fields than the CMIP3 archive did.

After the original GCM files were obtained from their host archive, they were subjected to a basic quality control to ensure the files contained the days and the data they stated that they did, that the data was within reasonable bounds for the variable listed, and that any missing data were identified. This analysis showed that many original GCM files had errors or peculiarities that were catalogued by this project before conducting the downscaling.

No attempt was made to select a sub-set of GCMs that performed better than others, as previous literature has showed that it is difficult, if not impossible, to identify such a sub-set for the continental U.S. (e.g. Knutti, 2010; Randall et al. 2007) However, the bias and error analysis conducted in a previous project identified one CMIP3 model (BCCR-BCM2) with consistently poor performance. One CMIP5 model (FGOALS-s2) was also withdrawn from the archive in 2013. As a result, simulations from these two GCMs were removed from the dataset generated by this project.

**Table 1.** CMIP3 and CMIP5 global climate modeling groups and their models used in this analysis. Those marked with (\*) archived only time slices: 1961-2000, 2046-2065 and 2081-2100. Those marked with (^) begin in 1961 (for CMIP3) and begin in 1960 and end in 2099 (for CMIP5). Those marked with a (†) have only 360 days per year. All other models archived full daily time series from 1960 to 2099 (CMIP3) and 1950 to 2100 (CMIP5).

Origin	CMIP3 model(s)	CMIP3 scenarios	CMIP5 model(s)	CMIP5 scenario(s)
Beijing Climate Center Climate System Model	N/A	N/A	bcc-csm1-1-m	4.5, 8.5
National Center for Atmospheric Research, USA	CCSM3 PCM	A1FI, A2, A1B, B1 A1FI, A2, A1B, B1	CCSM4	4.5, 8.5
Canadian Centre for Climate Modelling and Analysis, Canada	CGCM3.1 – T47 <sup>^</sup> CGCM3.1 – T63 <sup>^</sup>	A2, A1B, B1 A2, A1B, B1	N/A	N/A
Centro Euro-Mediterraneo per I Cambiamenti Climatici	N/A	N/A	CMCC-CM CMCC-CMS	4.5, 8.5 8.5
Centre National de Recherches Meteorologiques, France	CNRM-CM3	A2, A1B, B1	CNRM-CM5	4.5, 8.5
Commonwealth Scientific and Industrial Research Organisation, Australia	CSIRO-Mk3.0* <sup>^</sup>	A2, A1B, B1	CSIRO-MK3.6.0 ACCESS 1-0 ACCESS 1.3	4.5, 8.5 4.5, 8.5 4.5, 8.5
Max Planck Institute for Meteorology, Germany	ECHAM5/MPI	A2, A1B, B1	MPI-ESM-LR MPI-ESM-MR	4.5, 8.5 4.5, 8.5
National Institute of Meteorological Research/Korea Meteorological Administration	ECHO-G (with MPI) <sup>†</sup>	A2, A1B, B1	N/A	
NOAA Geophysical Fluid Dynamics Laboratory, USA	GFDL CM2.0 GFDL CM2.1	A1FI, A2, A1B, B1 A2, B1	GFDL-CM3	8.5
NASA Goddard Institute for Space Studies, USA	GISS-AOM* <sup>^</sup>	A1B, B1	N/A	
UK Meteorological Office Hadley Centre	HadCM3 <sup>†</sup> HadGEM1 <sup>†</sup>	A1FI, A2, A1B, B1 A2, A1B	HadGEM2-CC <sup>^†</sup>	4.5, 8.5
Institute for Numerical Mathematics, Russian	N/A	N/A	INMCM4	4.5, 8.5
Institut Pierre Simon Laplace, France	N/A	N/A	IPSL-CM5A-LR	4.5, 8.5
Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, and National Institute for Environmental Studies, Japan	MIROC3.2 (hires)* <sup>^</sup> MIROC3.2 (medres)* <sup>^</sup>	A1B, B1 A2, A1B, B1	MIROC5	4.5, 8.5
Meteorological Research Institute, Japan	MRI-CGCM2.3.2* <sup>^</sup>	A2, A1B, B1	MRI-CGCM3	4.5, 8.5

The CMIP5 GCM simulations used in this project consist of 16 model outputs archived by the Earth System Grid with continuous daily maximum and minimum temperature and precipitation outputs available for historical and the RCP 8.5 future scenario and 14 available for historical and the RCP 4.5 future scenario. No additional simulations were obtained from individual modeling group archives. The full list of CMIP5 GCMs used, their origin, the scenarios available for each, and the time periods covered by their output are given in Table 1.

### Downscaling Method

**This project used the statistical Asynchronous Regional Regression Model (ARRM). It was selected because it is able to resolve the tails of the distribution of daily temperature and precipitation to a greater extent than the more commonly used Delta and BCSD methods, but is less time-intensive and therefore able to generate more outputs as compared to a high-resolution regional climate model.**

Dynamical and statistical downscaling represent two complimentary ways to incorporate higher-resolution information into GCM simulations in order to obtain local to regional-scale climate projections. Dynamical downscaling, often referred to as regional climate modeling, uses a limited-area, high-resolution model to simulate physical climate processes at the regional scale, with grid cells typically ranging from 10 to 50km per side. Statistical downscaling models capture historical relationships between large-scale weather features and local climate, and use these to translate future projections down to the scale of any observations—here, both individual weather stations as well as a regular grid.

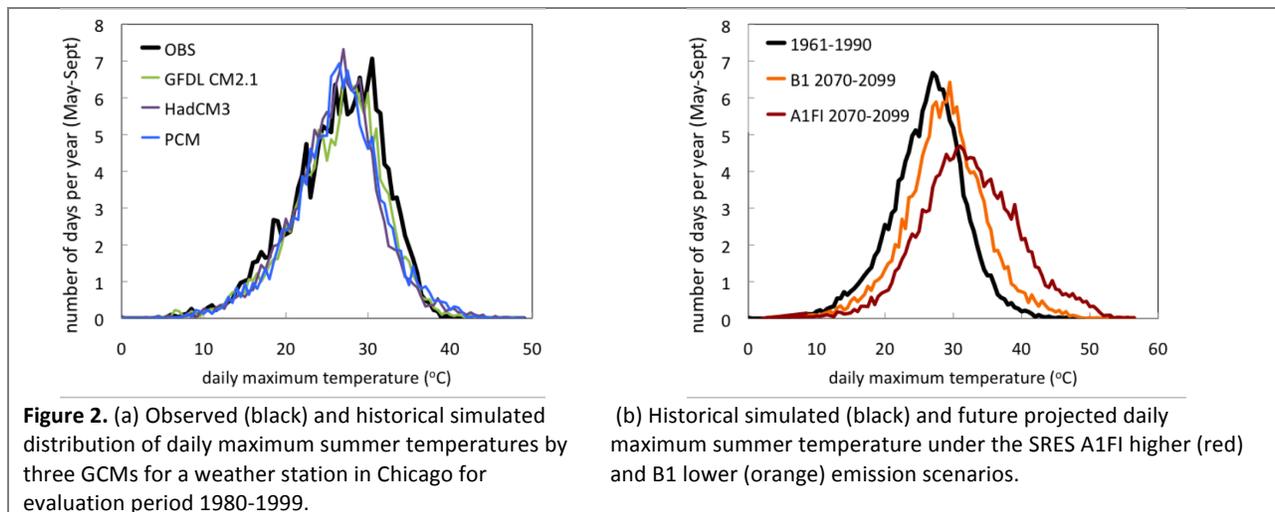
Statistical models are generally flexible and less computationally demanding than regional climate models, able to use a broad range of GCM inputs to simulate future changes in temperature and precipitation for a continuous period covering more than a century. Hence, statistical downscaling models are best suited for analyses that require a range of future projections that reflect the uncertainty in future scenarios and climate sensitivity, at the scale of observations that may already be used for planning purposes. If the study is more of a sensitivity analysis, where using one or two future simulations is not a limitation, or if it requires multiple surface and upper-air climate variables as input (and has a generous budget!), then regional climate modeling may be more appropriate.

Each commonly used downscaling method has its own benefits, and each can be sufficient for certain applications. For example, the simple delta or “change factor” approach does a good job with downscaling annual or seasonal mean temperature (as demonstrated in Figures 3 and 4). Regional climate models require large amounts of computing power, but provide consistent high-resolution projections for a broad range of surface and upper-air variables. None of these existing methods, however, allow for using multiple climate models and scenarios as input while downscaling to any spatial scale (including both station-based and gridded), and adequately resolving projected changes in daily climate extremes, at the same time.

For that reason, in this project we used a relatively new statistical downscaling model, the Asynchronous Regional Regression Model, or ARRM (Stoner et al., 2012). ARRM uses asynchronous quantile regression, originally developed by Koenker and Bassett (1978) to estimate conditional quantiles of the response variable in econometrics. Dettinger et al. (2004) was the first to apply this statistical technique to climate projections to examine

simulated hydrologic responses to climate variations and change, as well as to heat-related impacts on health (Hayhoe et al., 2004).

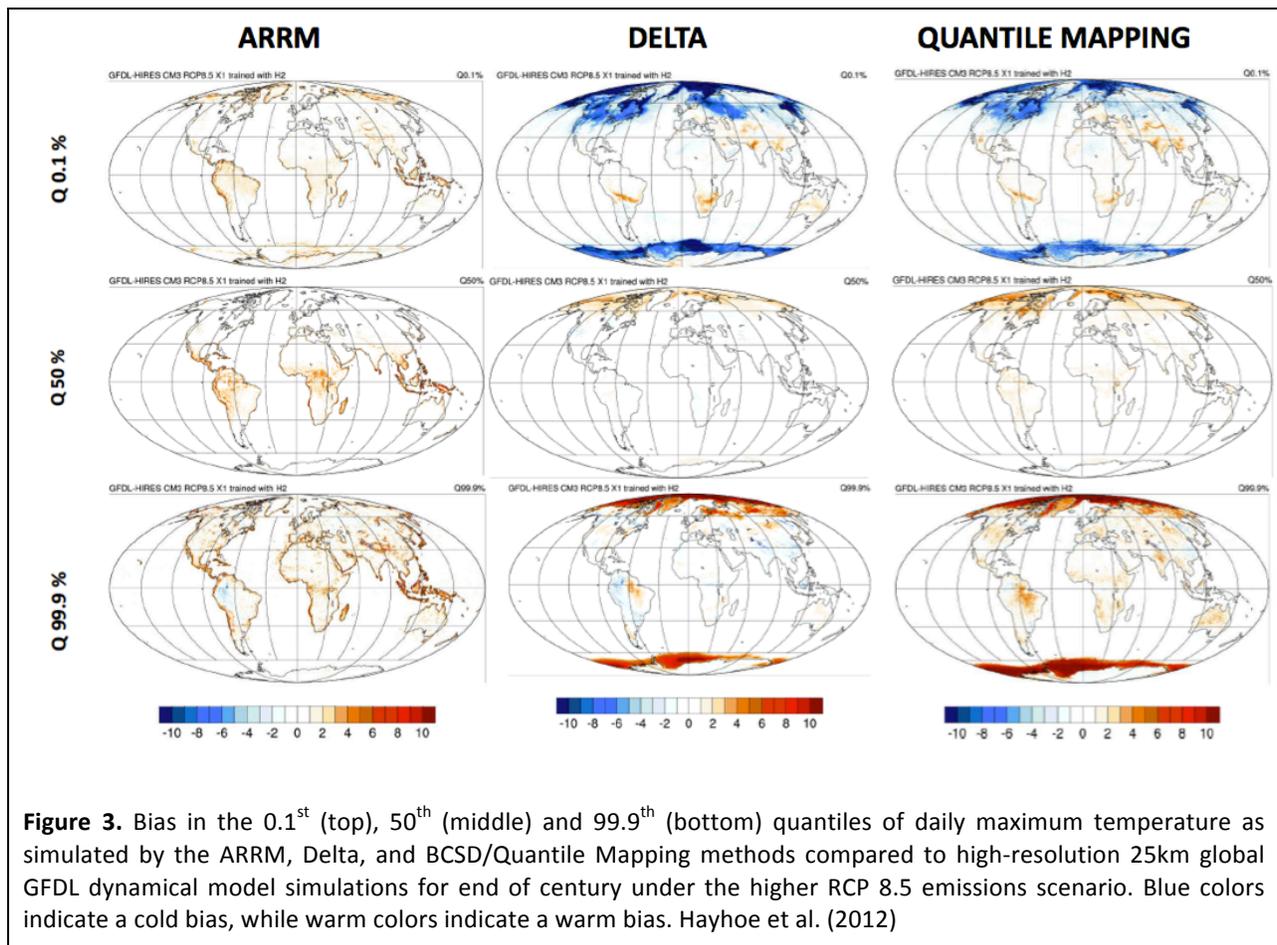
ARRM expands on these original applications with modifications specifically aimed at improving the ability of the model to simulate the shape of the distribution including the tails, including use of a piecewise rather than linear regression to accurately capture the often non-linear relationship between modeled and observed quantiles, and bias correction at the tails of the distribution. It is a flexible and computationally efficient statistical model that can downscale station-based or gridded daily values of any variable that can be transformed into an approximately symmetric distribution and for which a large-scale predictor exists. A quantile regression model is derived for each individual grid cell or weather station that transforms historical model simulations into a probability distribution that closely resembles historical observations (Figure 2a). This model can then be used to transform future model simulations into distributions similar to those observed (Figure 2b). More information on the ARRM method is provided in **APPENDIX D**, “An asynchronous regional regression model for statistical downscaling of daily climate variables” by Stoner et al. (2012).



Both statistical and dynamical downscaling models are based on a number of assumptions, some shared, some unique to each method. Two important shared assumptions are the following: first, that the inputs received from GCMs are reasonable, i.e. that they adequately capture the large-scale circulation of the atmosphere and ocean at the skillful scale of the global model; and second, that the information from the GCM fully incorporates the climate change signal over that region. This first assumption was evaluated in this project (see **ANALYSIS and RESULTS**).

All statistical models are based on a crucial assumption often referred to as **stationarity**. Stationarity assumes that the relationship between large-scale weather systems and local climate will remain constant over time. This assumption may be valid for lesser amounts of change, but could lead to biases under larger amounts of climate change (Vrac et al., 2007).

In a separate USGS-funded TTU-GFDL project, we are currently evaluating the stationarity of three downscaling methods, including the ARRM method (used here), the delta approach, which adds a “delta” derived from GCM output to observed mean annual, seasonal, or monthly values in order to get future values (e.g., Hay et al., 2000; as used in USGCRP, 2000); and the Bias Correction-Statistical Downscaling (BCSD) model, which uses a quantile mapping approach to downscale monthly AOGCM-based temperature and precipitation to a regular grid (Wood et al. 2004; as used in Hayhoe et al. 2004, 2008 and USGCRP, 2009; available from ClimateWizard.org and the DOE Green Data Portal). In this ongoing project, high-resolution 25km GFDL global model simulations for end-of-century under the higher RCP 8.5 scenario have been coarsened and used as input to these three statistical downscaling methods. The resulting projections are then compared directly to the high-resolution output to determine the extent to which the assumption of stationarity holds true. Where biases are small, stationarity is a reasonable assumption. Where biases are large, the assumption of stationarity fails.



To examine the stationarity in daily maximum temperatures, Figure 3 compares biases from the 0.1th to the 99.9th quantile (i.e. from the coldest day in 1000 to the hottest day in 1000). These biases represent the difference between daily maximum temperature values

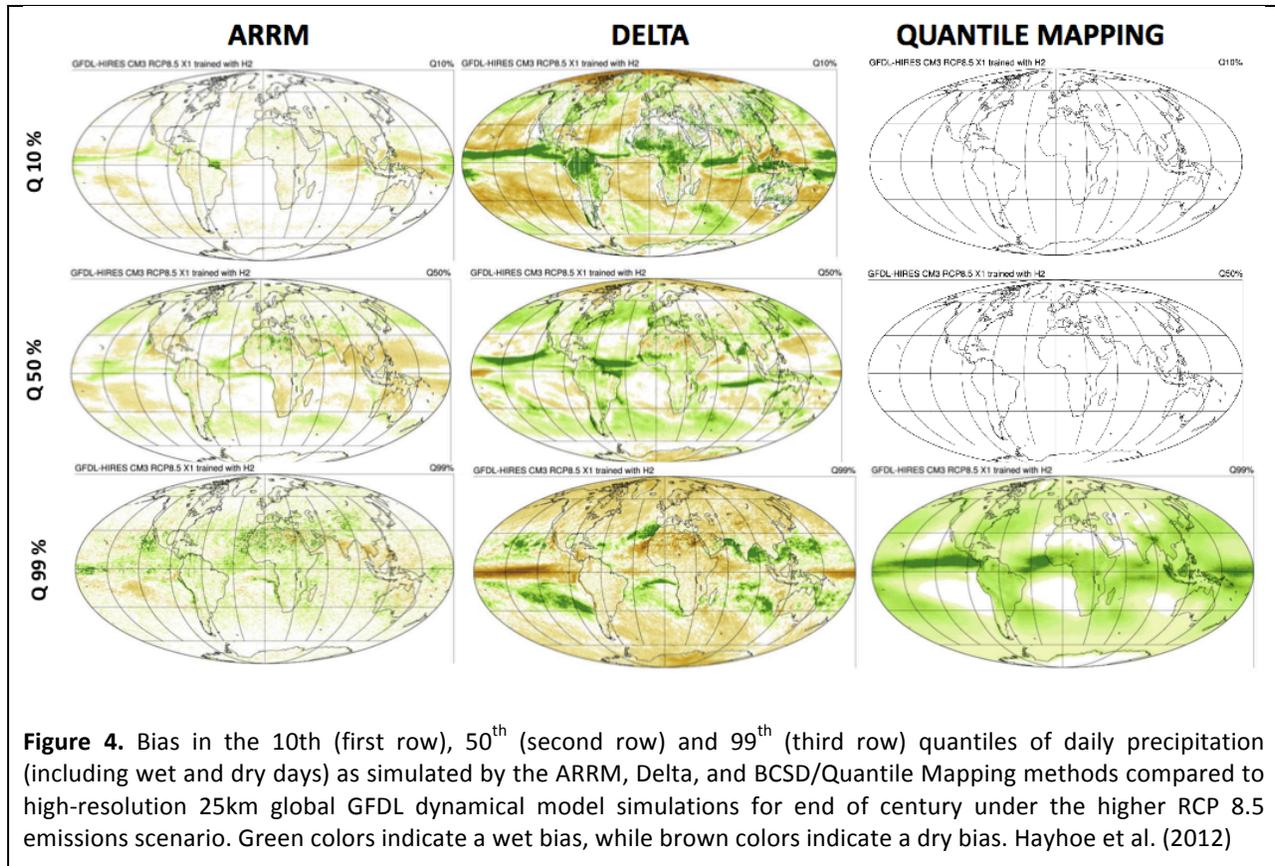
simulated by the 25km global climate model vs. those simulated by each statistical model, using coarsened global climate model fields as input.

Preliminary analyses show that the assumption of stationarity holds true over much of the world for the lower and middle of the distribution. The delta and BCSD methods exhibit a strong cold bias of up to 10°C at higher latitudes for cold temperatures. The BCSD method also begins to exhibit a systematic warm bias across the central U.S. at high temperatures. The only location where ARRM performance is systematically non-stationary is at high temperatures (at and above the 99.9<sup>th</sup> quantile) *along coastal areas*, with warm biases up to 6°C. This may be due to the statistical model's inability to capture dynamical changes in the strength of the land-sea breeze as the temperature differences between land and ocean are exacerbated under climate change; the origins of this feature are currently under investigation.

This bias has important implications for the climate projections generated for Puerto Rico and the Caribbean, since most station locations in this region are coastal. It suggests that estimated changes in days hotter than the 1-in-100 hottest historical day (e.g. the historical ~3 to 4 hottest days of the year) may be subject to temperature biases that increase in magnitude such that biases for the 1-in-1000 hottest days (e.g. the hottest day in 3 years) may be as large as the projected changes in the temperature of those days by end-of-century under a higher emissions scenario.

For precipitation, Figure 4 compares biases from the 10<sup>th</sup> to the 99.9<sup>th</sup> quantile for wet days only. Green colors are used to indicate where the statistical models over-estimate precipitation relative to the global model, while brown colors show where they underestimate future precipitation. Here, the BCSD approach, originally designed for forecasting streamflow that typically depends on accumulated precipitation over timescales of weeks (i.e. not high quantile events), shows a remarkable near-zero bias up to the 90<sup>th</sup> quantile. After that point, however, it rapidly develops a systematic positive bias in precipitation covering almost the entire world. The Delta method exhibits a systematic positive (wet) bias for low precipitation over land that shifts to a systematic negative (dry) bias for high precipitation over land. The ARRM method is characterized by a spatially variable bias at all quantiles that *is generally not systematic*, and varies from approximately -30 to +30% depending on location.

Although the downscaling model is purposely designed to be applicable to any variable with a relatively symmetric distribution, predictors must be pre-selected for each variable and there are some differences in the initial processing of each predictor that can improve the performance of the model in downscaling. The ARRM method has been specifically designed to allow for user-selected predictors. For the sake of consistency, however, in this project predictors were chosen to be the same variables as the predictands: 2m maximum and minimum temperature and 24h cumulative precipitation. These are the most frequently-archived daily output from both CMIP3 and CMIP5 GCMs.



### Gridded Observations and Reanalysis

This project used gridded observations of monthly precipitation (GPCP), sea surface temperature (NOAA), mixed layer depth (Levitus) and atmospheric surface and upper-air fields (ERA-40). Gridded fields were used to identify large-scale drivers of temperature and precipitation variability over the Caribbean and evaluate global climate model performance relative to observed.

Gridded precipitation data was obtained from the global merged precipitation data of the Global Precipitation Climatology Project, GPCP version 2 (Adler et al. 2003). As stated on the project website (<http://www.gewex.org/gpcp.html>), the GPCP combines data from 6,000 rain gauge stations, satellite geostationary passive microwave instruments, and sounding observations. This dataset is described as the most complete analysis of rainfall available over the global oceans, as well as adding additional detail over land areas with sparse station records, such as the Caribbean. GPCP data is gridded to 2.5 degrees at monthly resolution from 1979 to 2008 (data available online at: <ftp://precip.gsgc.nasa.gov/pub/gpcp-v2.1/psg>).

Upper-air and surface atmospheric fields, including monthly mean geopotential height, zonal and meridional wind, vertical velocity, specific humidity, and 2 meter temperature was obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) 40 year reanalysis, ERA-40 (Uppala et al. 2005; data available online at: <http://dss.ucar.edu/datasets/ds119.0>). ERA-40 output is available at a resolution of approximately 1.4 degrees for the period from September 1957 to August 2002.

Reanalysis does not consist of actual observations, but rather of output from a numerical weather simulation model constrained to match available observations at regular intervals. In broad terms, reanalysis can be viewed as informed dynamical interpolation between observations that, for regions such as the Caribbean, are often relatively sparse.

Sea surface temperature (SST) data used here comes from the National Oceanic & Atmospheric Administration (NOAA) Optimum Interpolation (OI) SST V2 (Reynolds et al. 2002; data available at:

<http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>). The SST data has a spatial resolution of 1 degree with monthly mean data available for 27 years, from 1982 to 2008.

Monthly mean ocean mixed layer depth was obtained from the Levitus World Ocean Atlas (Monterey & Levitus, 1997; data available at:

<http://www.esrl.noaa.gov/psd/data/gridded/data.nodc.woa94.html>). The mixed layer depth data was available at a spatial resolution of 0.25 degrees, from 1900 to 1992.

### Station Observations

**This project used long-term station data from the Global Historical Climatology Network, supplemented with additional station data from MIDAS and NOAA/NCDC. All station data was quality-controlled to remove questionable data points before being used to train the statistical downscaling model.**

Long-term weather station records were obtained from the Global Historical Climatology Network (GHCN; <http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>) and supplemented with additional records from the U.K. Met Office Integrated Data Archive System Land and Marine Surface Stations Data (MIDAS, <http://badc.nerc.ac.uk/data/ukmo-midas/>) and the National Climatic Data Center cooperative observer program (NCDC-COOP, <http://www.ncdc.noaa.gov/land-based-station-data/cooperative-observer-network-coop>).

MIDAS was not available in daily format; thus, three-hour instantaneous temperature data was extracted and fit to a diurnal function and cumulative precipitation was extracted and summed before being evaluated relative to WMO standards (e.g. maximum daily temperature must occur between 6am and 6pm; minimum nighttime temperature between 6pm and 6am; precipitation is summed from 0h to 21h local time).

To train the downscaling model, the observed record must be of adequate length and quality. To appropriately sample from the range of natural climate variability at most of the station locations, and to produce robust results without over-fitting, stations were required to have a minimum of 20 consecutive years of daily observations overlapping GCM outputs with less than 50% missing data after quality control. When these limits were applied, the number of usable stations for the Caribbean totaled 284 for maximum temperature, 282 for minimum temperature, and 231 for precipitation. The latitude, longitude, and station names of the weather stations for which downscaled projections were generated are provided in **APPENDIX A**.

All datasets were incorporated into the quality control framework described in **ANALYSIS AND FINDINGS**. Long-term records from these datasets were then downscaled and compiled into a database of high-resolution projections for the Caribbean described in **PROJECT RESULTS**.

## Computational Methods

Data file formats and computational approaches used represent the standard for climate model data and analysis. Datasets are archived as ascii text files (observed station data) and self-describing netCDF files (gridded observations and station-based climate projections). All codes were written using the statistical programming language R (<http://cran.us.r-project.org/>). Bar charts were prepared using Microsoft Excel and maps using the NCAR Command Language (<http://www.ncl.ucar.edu/>) and ArcGIS (<http://www.esri.com/software/arcgis>) and stored in postscript or Adobe PDF format.

## PROJECT RESULTS

This project produced three specific products: a quantitative dataset of raw climate projections for long-term weather stations in the Caribbean; a qualitative analysis of global climate model performance over the Caribbean; and a quantitative analysis of projected climate changes for Puerto Rico stations, as follows:

**ONE. A database of daily projections for 284 (maximum temperature), 282 (minimum temperature) and 231 (precipitation) individual long-term weather stations in the Caribbean.**

This dataset consists of daily maximum and minimum temperature and 24h cumulative precipitation for each day from 1960 to 2099 (CMIP3) and 1950 to 2100 (CMIP5) for which global climate model output was available. When GCM output was not available for a given day, or was clearly erroneous, the value for that day was replaced with an “NA” value. NA values in the downscaled projections for each weather station are not errors; they are merely indicative of a lack of input information for that day.

The majority of global climate model fields used in this analysis were obtained from the WCRP CMIP3 and CMIP5 archives, which maintains standards of data provision and quality control. (SRES A1FI scenarios were not archived by CMIP3: these simulations were obtained directly from individual modeling groups.) Before the global model fields could be used, they were quality-controlled for specific problems including:

1. Missing data (which could range from a few random days to entire decades in the middle of a simulation)
2. Incorrect values (unrealistic data points far above or below historical observed maxima or minima)
3. Mis-labeled files (files that did not contain the variable or data listed in the file name and header)

In addition to these errors, each model also had its own peculiarities that had to be standardized before the predictors could be incorporated into the downscaling model: days per year ranging from 360 to 366 (for models with 360 days, “NA” values were added for the 31<sup>st</sup> of each month from August to December), different data units, and limited information (e.g., five CMIP3 models only had limited future time slices available, others began in 1961 instead of 1960).

These and other known quality issues and model peculiarities are listed in Table 3.

**Table 3.** Known quality control problems with original global climate model outputs and model peculiarities that were documented and, if possible, standardized before the model was used in Step One.

Global climate model	Known issues
<b>CMIP5</b>	
CMCC-CMS	RCP 4.5 not available
GFDL-CM3	RCP 4.5 not available
HadGEM2-CC	Model has 360 days per year. These were divided up such that the last days of May, July, Aug, Oct and Dec are <b>always missing</b> . Years 1950-1959 and 2100 are also missing.
CCSM3	Temperature data for historic and future periods not initialized from the same run, creating disconnect between the two files if one were to plot a continuous time series joining the 2 files. This situation is unavoidable for CCSM, as most of its historical simulations have erroneous tmax/tmin values (they were accidentally overwritten with the variable TREFHT by the original modeling group).  A1fi Tx/Tn: b30.030h to b30.099a (no match) A1fi Pr: b30.030a to b30.099a (matched)  A1B Tx/Tn: b30.030h to b30.042g (no match) A1B Pr: b30.040a to b30.030a (matched)  A2 Tx/Tn: b30.030h to b30.042e (no match) A2 Pr: b30.030e to b30.042e (matched)
CGCM3-T47	A1fi scenario not available. Simulation begins in 1961. 1960 is missing.
CGCM3-T63	A1fi scenario not available. Simulation begins in 1961. 1960 is missing.
CNRM	A1fi scenario not available. Leap years removed before downscaling.
CSIRO	A1fi scenario not available. Future data not available for 2000-2045 and 2066-2080.
ECHAM5	A1fi scenario not available. Leap years removed before downscaling.
ECHO	A1fi scenario not available. Model has 360 days per year. These were divided up such that the last days of May, July, Aug, Oct and Dec are <b>always missing</b> .
GISS-AOM	A1fi and A2 scenarios not available. Future data not available for 2000-2045 and 2066-2080.
GFDL CM2.0	A1fi and A1B scenarios not available.
GFDL CM2.1	No known issues.
HadCM3	Model has 360 days per year. These were divided up such that the last days of May, July, Aug, Oct and Dec are <b>always missing</b> . For the A1B simulation, the last 30 days are missing (Dec 2099) and 10 years between 2080-2089 are missing for precipitation. For A2 and B1, the last 60 days of the historical period (Nov-Dec 1999) are

	missing and the first 10 days of Nov 2000 are missing. For B1, the year 2000 is missing.
HadGEM	Model has 360 days per year. These were divided up such that the last days of May, July, Aug, Oct and Dec are <b>always missing</b> . A1fi and B1 scenarios not available.
Miroc-med	A1fi scenarios not available. Future data not available for 2000-2045 and 2066-2080. Leap years removed before downscaling.
Miroc-hi	A1fi and A2 scenarios not available. Future data not available for 2000-2045 and 2066-2080. Leap years removed before downscaling.
MRI_CGCM2	A1fi scenario not available. Future data not available for 2000-2045 and 2066-2080.
PCM	PCM A2, A1B and B1 scenarios are based on B06.08 to 1980 and B07.08 from 1980 to 2099 but daily data for B06.08 is not available before 1980 so B06.57 was used instead. There will be a level shift at 1980 because of this. The last 61 days of 2099 (Nov-Dec) are also missing in the A1B scenario.

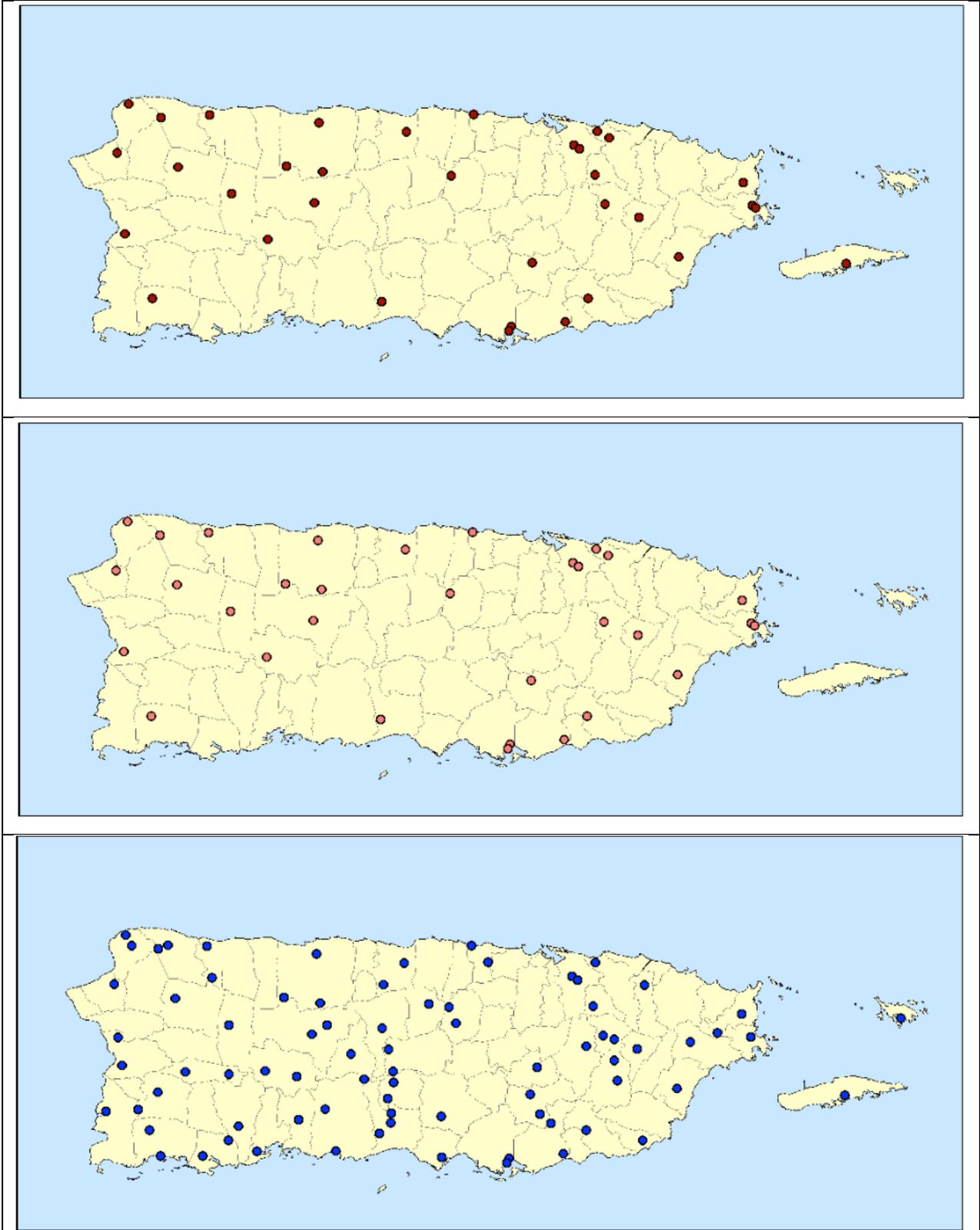
Although most of the station data used in this project was retrieved from datasets (GHCN, MIDAS) that have been nominally quality-controlled, previous encounters with instances of clearly erroneous values (e.g. very cold days in warm locations, days where maximum temperature was lower than minimum temperature, days with negative precipitation) in the observations that strongly degrade the quality of the statistical downscaling model has led us to construct our own quality control process for filtering of station data.

Assembling all station data, we first created a master file documenting the name, latitude, and longitude of every station in the various datasets available. Using latitude and longitude values, stations within 1km of each other were assumed to be co-located and only the longest unique station record for each variable was retained. As the station data originated from several different datasets, this was a helpful step to remove potential duplicate entries.

The quality control process consists of two steps: first, individual quality control for each station; and second, a nearest neighbor approach to validate outliers identified relative to the climatology of each month. Individual quality control identified and replaced with N/A any values that failed the following tests:

- **Tmin>Tmax** – Daily minimum temperature exceeds the daily maximum temperature.
- **Repeated Values** – Daily maximum temperature, daily minimum temperature, daily average temperature, or non-zero daily precipitation values repeat for 5 or more consecutive days to within one tenth of a mm per day or one hundredth of a degree Celsius.

In the second step of the quality control process, the merged database of station locations is first used to identify up to 10 “nearest neighbors” for each individual weather station



**Figure 5.** Long-term weather stations with sufficient data to use in generating station-specific projections for Puerto Rico: maximum temperature (top), minimum temperature (middle), and precipitation (bottom)

within 100km of its location. For each weather station, the monthly (for temperature) and seasonal (for precipitation) distributions are ranked and the highest and lowest N values are identified. The nearest neighbor stations are then queried to see if the days on which values 1 through N occur are also days in which the highest M values for the neighbor station occur, plus or minus one day on either side to account for weather systems which may be moving through the area close to midnight. Here, N is taken to be 100 and M to be 500. If any value of N does not occur within +/- one day of a value M for any neighboring station, the value of N is replaced with an NA.

The resulting files are then scanned to identify any stations with less than 3650 real values and less than 200 values for any given month. After the quality control and filtering process was complete, a total of 284 (maximum temperature), 282 (minimum temperature) and 231 (precipitation) Caribbean stations were available to be downscaled using the ARRM model described previously and the GCM inputs listed in Table 1. The locations of the stations identified for Puerto Rico are mapped in Figure 5.

The resulting station-level projections were archived in ASCII files for individual locations, with each file identified by latitude, longitude, and station ID. There is one file for each station location/variable/CMIP combination. Each file contains N+3 columns, where N is the number of model/scenario combinations for the given set of CMIP data. The first 3 columns of each file contain the year, month, and day corresponding to each row. There are a total of 51,100 days per CMIP3 file and 55,115 days per CMIP5 file.

*As there are a large number of these files, they are provided in electronic format only.*

## **TWO. An assessment of CMIP3 and CMIP5 global model ability to reproduce the large-scale atmosphere and ocean drivers of observed variability in temperature and precipitation over the Caribbean.**

Models were evaluated according to the criteria described in **ANALYSIS AND FINDINGS**. There was not a significant difference between models in terms of their ability to simulate changes in temperature; however, there was a distinct difference in model ability to simulate the drivers of precipitation. Through dynamical analyses, the models were divided up into 3 categories based on their ability to simulate large-scale atmospheric and oceanic features controlling precipitation in the Caribbean.

The results of the analysis of GCM ability to simulate drivers of precipitation are described in **APPENDIX B**, "Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations" by Jung-Hee Ryu and Katharine Hayhoe which has been accepted for publication in *Climate Dynamics*. A manuscript summarizing the results of the analysis of GCM ability to simulate drivers of temperature is currently in preparation; the main results of this analysis are summarized in the **ANALYSIS AND FINDINGS** section.

The final product of this analysis is a list of models in each of these groups, with the implication that models in group 1 (those able to reproduce the observed wet season and mid-summer drought) are better-able to simulate variability in precipitation than models in group 2 (those able to simulate a mid-summer drought and truncated wet season) or group 3 (those simulating a single wet season with no mid-summer drought).

*This list is given in **ANALYSIS AND FINDINGS**.*

### **THREE. Analysis of projected changes in 85 secondary climate indicators for Puerto Rico long-term weather stations.**

In addition to providing raw values of projected daily maximum and minimum temperature and precipitation, we also calculated annual values (140 values, one per year from 1960 to 2099, for CMIP3 simulations and 151 values, one per year for 1950 to 2100, for CMIP5 simulations) for 85 individual indicators of mean and extreme temperature and precipitation. The full list of these indicators is provided in Table 2 below.

Annual values were then averaged across stations by region (hot coastal vs. temperate inland, for temperature; dry northern coast, dry southern coast, and wet central region for precipitation), and across GCMs by group (good, fair, and poor), and across three future climatological 20-year time periods corresponding to global mean temperature increases of +1, 2, and 3°C relative to the 1971-2000 average.

This approach, averaging across GMT periods, was introduced by the 2011 National Research Council Report, *“Climate Stabilization Targets: Emissions, Concentrations, and Impacts over Decades to Millennia”* (NAS, 2011). A summary of this report, entitled *“Warming World: Impacts by Degree”* is available at: <http://nas-sites.org/americasclimatechoices/more-resources-on-climate-change/booklet-warming-world-impacts-by-degree/>. We use this approach as it has the advantage of comparing projected changes across a range of scenarios and climate sensitivities to resolve any differences in the magnitude and pattern of expected change *independent* of the uncertainty in either human scenarios or climate sensitivity. This approach also presents impacts within a policy-relevant framework, as national and international agreements (such as the EU target of +2°C) are more often couched in terms of global mean temperature or atmospheric concentration targets than in terms of a given emissions scenario.

Averages are calculated by first calculating annual average global average temperature for each model for each year from 1960 to 2099 (CMIP3) and 1950 to 2100 (CMIP5). Then, the 20-year running mean values were calculated for each year beginning with 1990-2009 and ending with 2080-2099 (CMIP3) and 2081-2100 (CMIP5). The first 20-year period in which global mean temperature was equal to or higher than +1/2/3°C compared to the 1971-2000 global mean temperature *of that simulation* was identified. This time period is different for each GCM: further into the future for models that simulate a climate system that is less sensitive to emissions from human activities, and nearer in time for models that simulate a more sensitive climate system. Finally, this date was used to select the 20-year period from the corresponding simulation to average for each indicator.

Yearly values of each indicator are provided in ASCII text files, one file for each station and indicator. Each file contains N+1 columns, where N is the number of model/scenario combinations for the given set of CMIP data. The first column of each file contains the year. There are a total of 140 years per CMIP3 file and 151 years per CMIP5 file.

Geographically-, model-based, and temporally-averaged indicators have been plotted and are available as images in **APPENDIX C** and in Excel files, one for each variable (maximum temperature, minimum temperature, and precipitation).

*The raw files of annual indicators are provided in electronic format only.*

**Table 2.** Primary and secondary climate indicators calculated for 3 global mean temperature targets (+1,2,3°C relative to 1971-2000). Wet season is defined as May to November. Dry season is defined as December to April. All changes are anomalies relative to historical 1971-2000 mean.

Indicator	Units
<b>ANNUAL AND SEASONAL AVERAGES</b>	
Average maximum temperature (wet season, dry season, and annual)	degrees C
Average maximum temperature (wet season, dry season, and annual)	"
Average cumulative precipitation (wet season, dry season, and annual)	mm per season or year
<b>ANNUAL AND SEASONAL EXTREMES</b>	
Number of warmer-than-historical maximum wet seasons	Number of years historical threshold exceeded (up to 20)
Number of colder-than-historical minimum years	"
Number of wetter-than-historical and drier-than-historical maximum years	"
<b>ANNUAL AND SEASONAL VARIANCE</b>	
Standard deviation of daily maximum temperature (wet season, dry, season, and annual)	degrees C
Same, for daily minimum temperature	"
Same, for cumulative precipitation, wet days only	log(mm)
Range in daily maximum temperature (wet season, dry, season, and annual)	degrees C
Same, for daily minimum temperature	"
Precipitation intensity, defined as cumulative precipitation divided by number of wet days (wet season, dry, season, and annual)	mm/day
<b>ANNUAL EXTREMES</b>	
Maximum temperature of 1, 3, 5, and 10-day hottest period	degrees C
Same, for coldest period (based on daily minimum temperature)	"
Same, for cumulative precipitation	mm
Total number of dry days defined as $pr < 0.01$ " (wet, dry, annual)	Days per year
Longest stretch or period of dry days, defined as $pr < 0.01$ " per day (wet, dry, annual)	Days
Average length of dry periods, defined as $pr < 0.01$ " per day (wet, dry, annual)	"
<b>THRESHOLDS</b>	
Daily maximum temperature above 80, 85, 90, 95 and 100°F	Days per year
Nighttime minimum temperature below 65 and 75°F	"
Nighttime minimum temperature above 80, 85 and 90°F	"
Number of wet days per year ( $pr > 0.01$ "	"
Heavy precipitation days ( $pr > 1, 2, 3$ inches)	"
<b>QUANTILES</b>	
Number of hot days per year with daily maximum temperature above the historical (1971-2000) 50, 75, 90, 99 and 99.9 <sup>th</sup> quantile	Average number of days per year
Number of cold nights per year with daily minimum temperature below the historical (1971-2000) 0.1, 1, 10, 25, and 50 <sup>th</sup> quantile	"
Number of wet days with precipitation exceeding the historical (1971-2000) 50, 75, 90, 99 and 99.9 <sup>th</sup> quantiles	"
<b>HYBRIDS</b>	
Number of hot, dry days per year (defined as precipitation $< 0.01$ " and daily maximum temperature $> 90$ <sup>th</sup> quantile	Average number of days per year
Number of cool, wet days per year (defined as precipitation $> 0.01$ " and daily maximum temperature $< 10$ <sup>th</sup> quantile	"

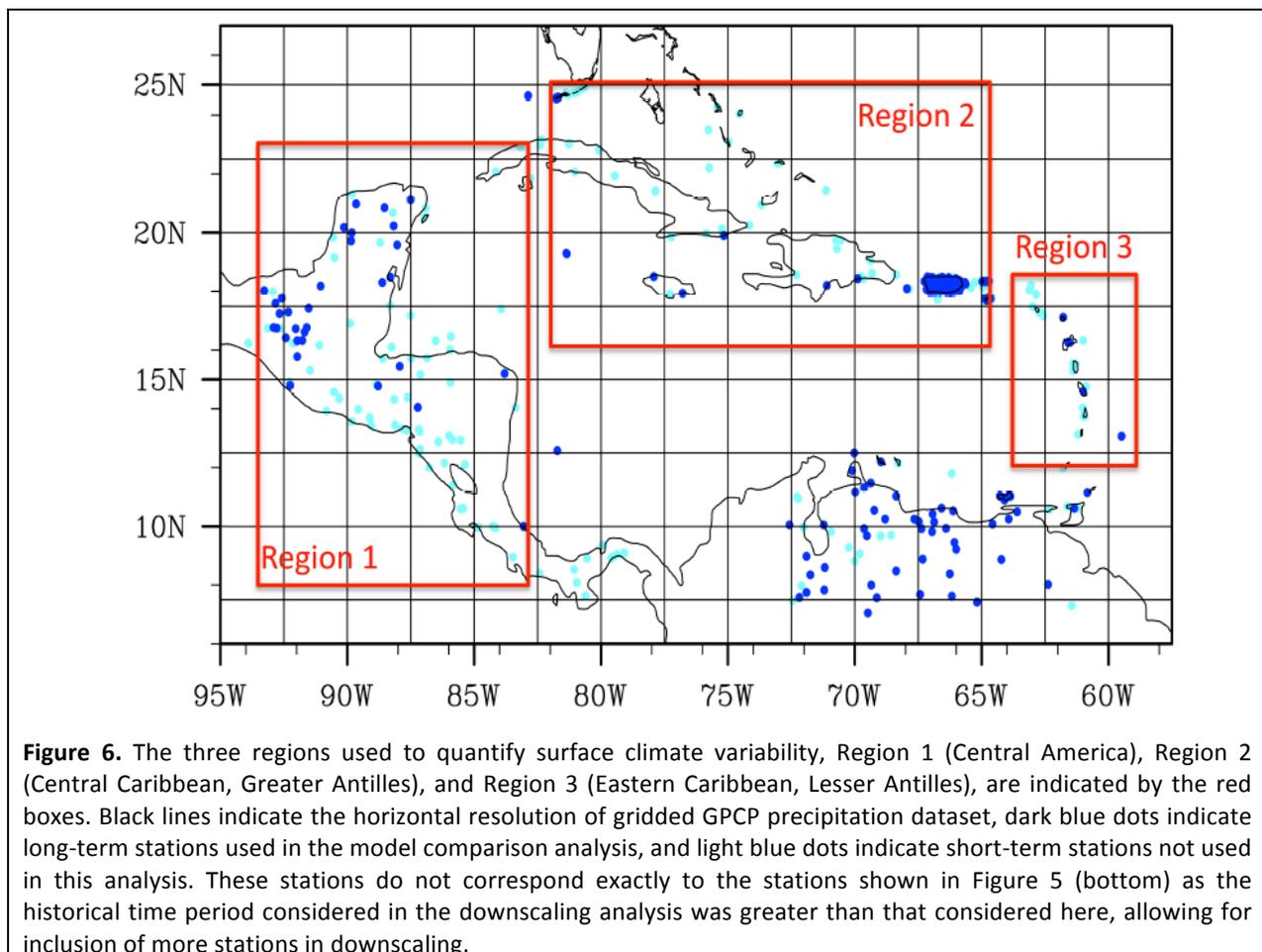
## ANALYSIS AND FINDINGS

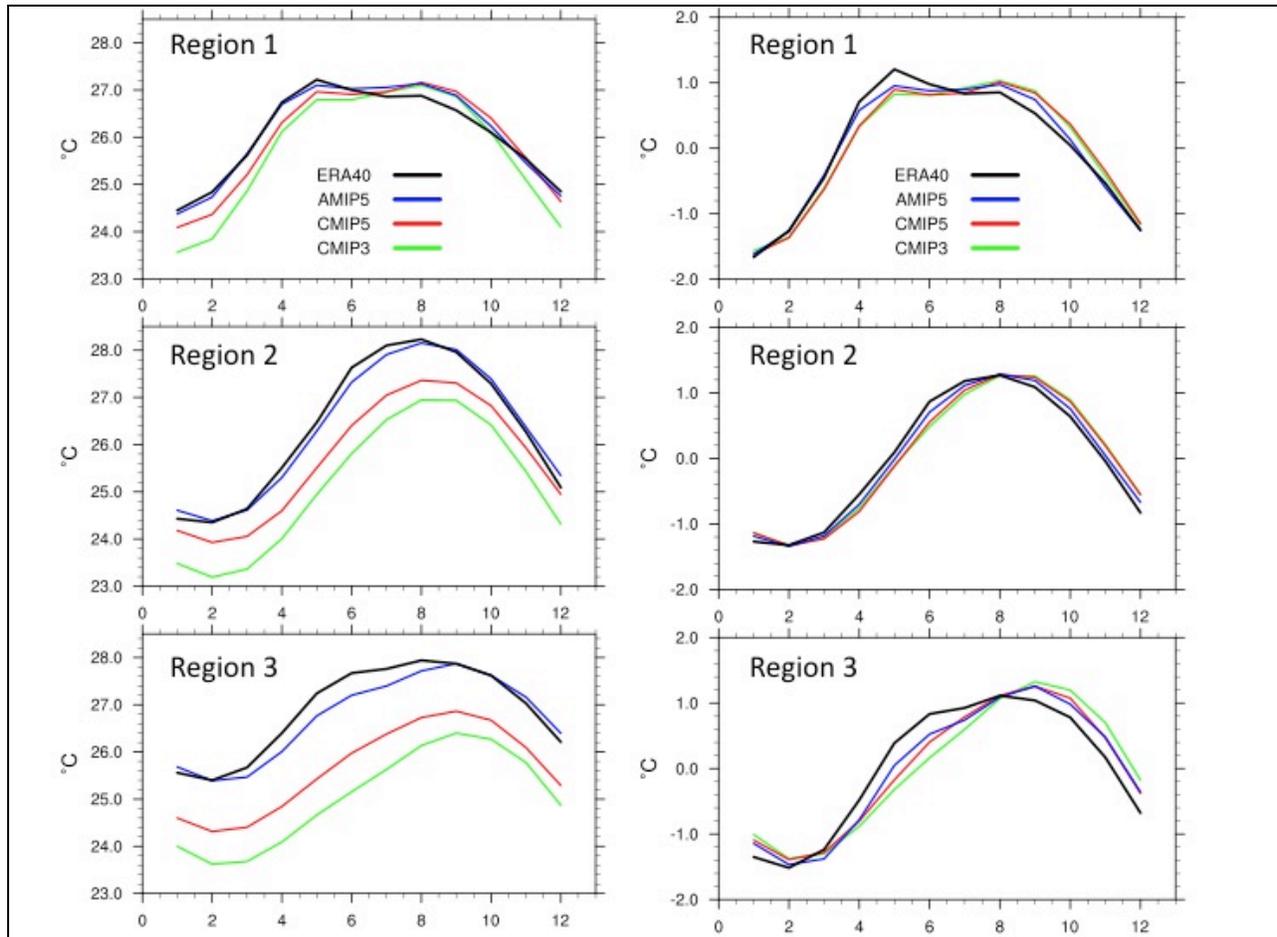
The previous section described the specific products generated by this project. This section focuses on the results of the analyses for product 2 (analysis of GCM ability to simulate large-scale drivers of Caribbean climate) and 3 (climate projections for Puerto Rico).

### Global Climate Model Performance: Caribbean Temperature

For temperature, almost every GCM showed a systematic lag in the seasonal cycle. Some improvement is seen in higher as compared to coarser resolution GCMs, for atmosphere-only (AMIP) models as compared to coupled atmosphere-ocean (CMIP) models, and for newer (CMIP5) as compared to older (CMIP3) models. Overall, however, this bias appears to be symptomatic of nearly every one of the 44 GCMs analyzed here.

To characterize temperature variability over the Caribbean, we first divided the Caribbean into three regions: Central America (east), the Greater Antilles (north-central), and Lesser Antilles (west). Figure 6 shows the regions on a map, superimposed on the resolution of the GPCP data and with dots indicating locations of long-term weather stations used in the analysis. For each region, we conducted a comprehensive literature review and analyzed composite maps of gridded observed and reanalysis-based ocean, surface, and upper-air fields to identify the large-scale atmospheric and oceanic features most relevant to temperature variability.



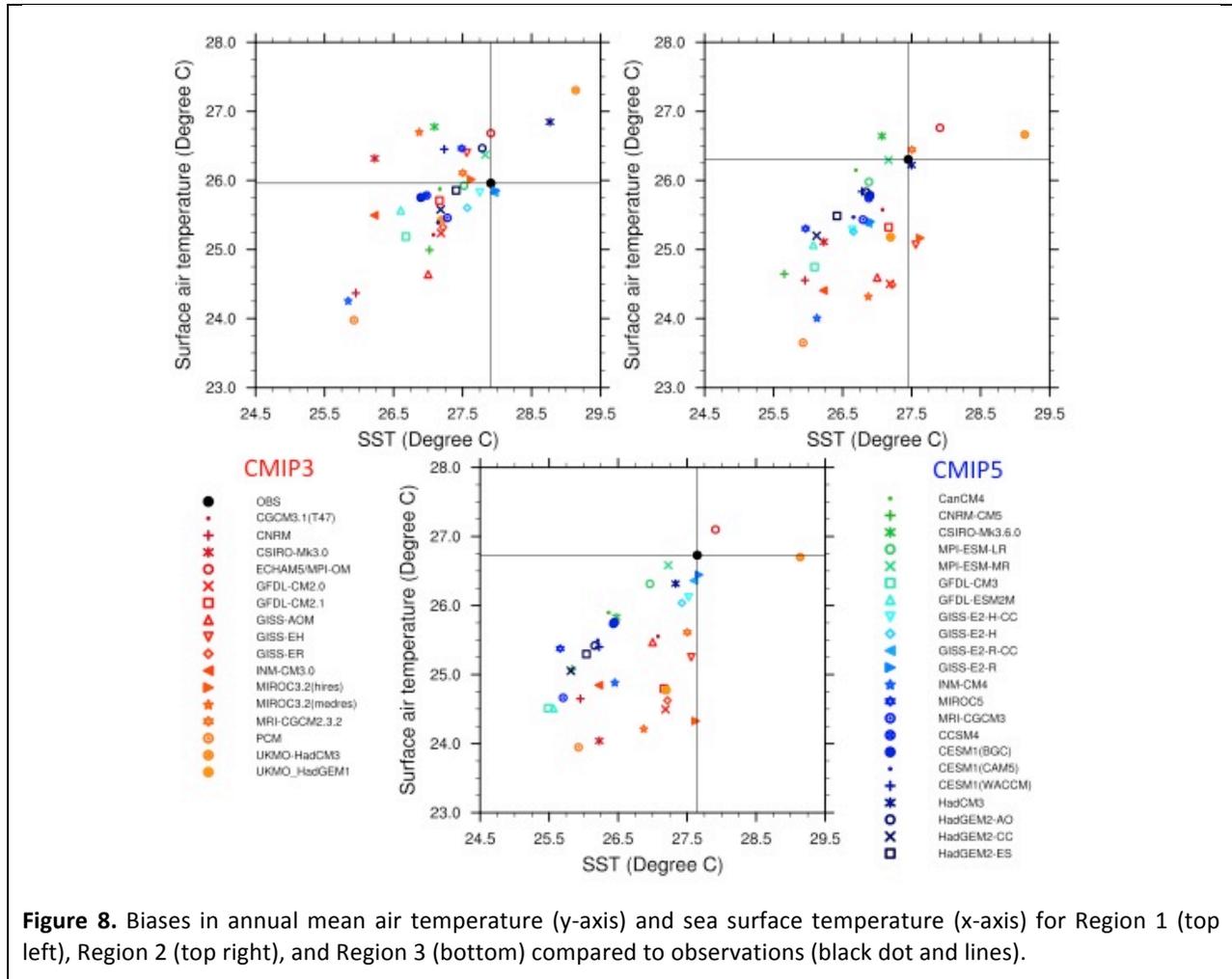


**Figure 7.** Raw (left) and normalized (right) annual cycles of temperature over the three Caribbean regions shown in Figure 6. Results have been averaged across all CMIP3 models (green), all CMIP5 models (red), and all AMIP atmosphere-only models (blue).

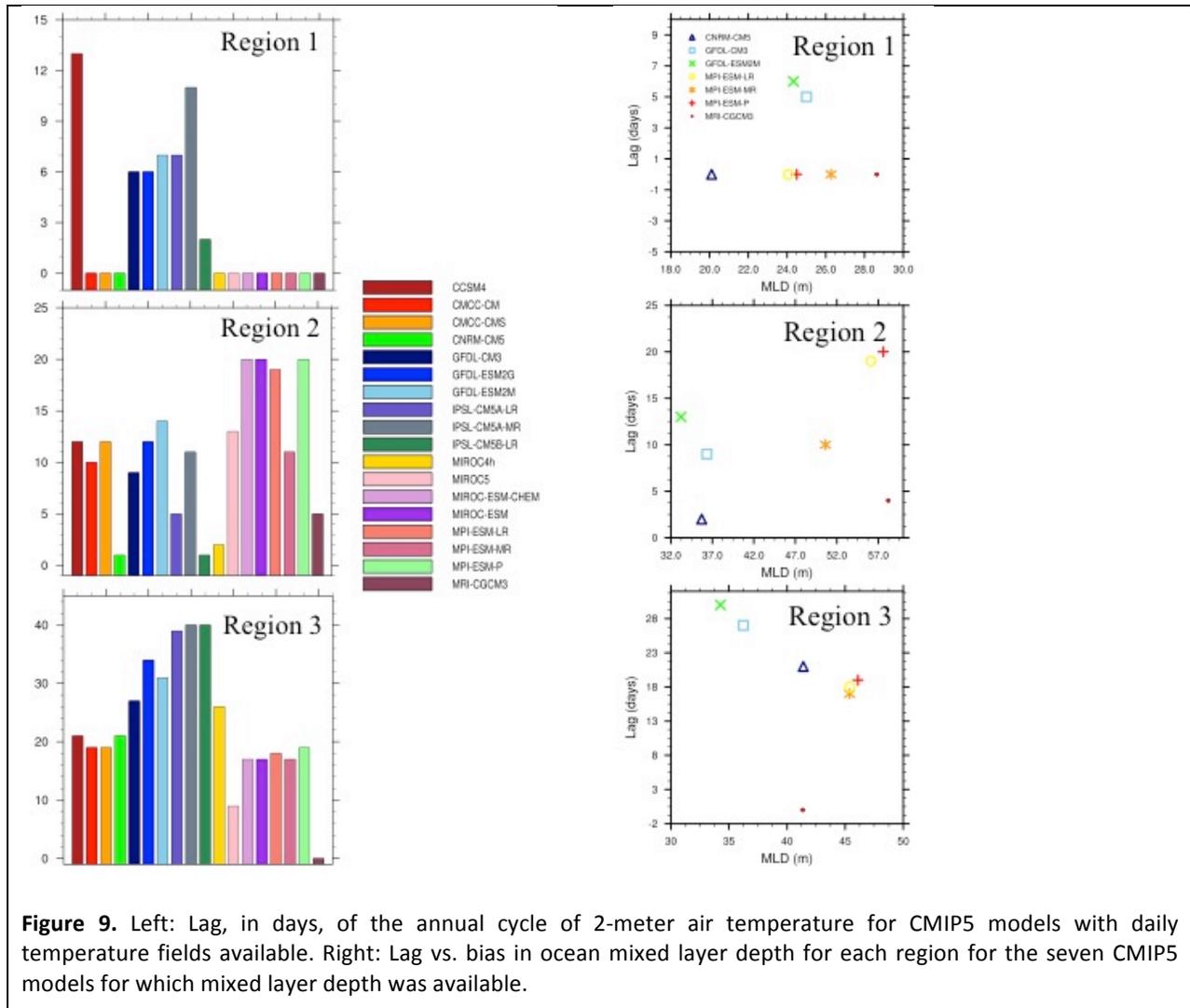
The first step in the analysis was to develop a climatology of annual temperature from available data. For each of the three regions, we constructed composite annual cycles of temperature using ERA-40 reanalysis output and compared the ERA-40 climatology with that derived from station data (station data not shown; ERA-40 based temperature is shown in Figure 7). Region 1 (Central America) is characterised by a nearly bimodal distribution, meaning that although the warmest month occurs in May, there is a second peak in temperature in August. Region 2 (Central Caribbean) has one warm season, with a peak in August. Region 3 (Eastern Caribbean) has the lowest amplitude of any annual cycle among the three regions, with a relatively longer and more persistent warm season beginning in spring and stretching well into fall.

The next step was to compare these observed climatologies with the annual cycles simulated by historical simulations from CMIP3 and CMIP5 GCMs as well as for AMIP (atmosphere-only) models, to eliminate potential biases in the temperature cycles arising from limitations in the ocean model or coupling between the ocean and atmosphere components of a coupled model. For consistency, we compared ERA-40 with GCMs as its

output fields for 2 meter temperature include similar proportions of land and ocean area to the GCM simulations (most are not precisely identical due to the different resolution of the various models, but all are more similar than comparing GCM outputs with land-based station records only). Climatological mean and standard deviations of the annual temperature cycle vary from one model to the next. These biases are relatively straightforward to correct when combining future projections with any type of downscaling technique. For that reason, we also normalized annual temperature cycle for each region for each of the GCM simulations and ERA-40 outputs. The resulting annual cycles are summarized by region in Figure 7.



For Region 1 (Central America), most GCMs simulate a bimodal distribution of temperature, but with an early peak that is too low and a later peak that is too high. For Regions 2 and 3, GCMs also are able to reproduce the shape of the distribution. For all three regions, however, there is a significant lag in GCM simulations: a lag that is smaller but not entirely eliminated in AMIP atmosphere-only versions of the CMIP models. The only CMIP3 GCM to correctly simulate the timing of onset in Region 2 is MRI, which includes flux adjustment. However, this model is also the only one to anticipate the end of the Region 2 warm season by about a month relative to ERA-40.



Comparing green (CMIP3) and red (CMIP5) simulations reveals a noticeable improvement in the bias in temperature values for Regions 2 and 3 (Figure 7, left) but not in the lag in the seasonal cycle (Figure 7, right). Nearly every model, however, is still too cool compared to observed air temperature. Comparing model biases in 2-meter air temperature with SST (Figure 8) reveals that biases in air temperatures closely track biases in SSTs.

For CMIP5 GCMs, we calculated the exact lag by day, to determine whether there was any correspondence between GCM spatial resolution and lag. Comparing one GCM with higher resolution with a similar GCM from the same family and lower resolution showed a very small improvement in lag (not shown) that was not evident when GCMs were compared by resolution across families. Rather, we found a stronger relationship between model *origin* and lag. Specifically, it was possible to group the GCMs into “families” characteristic lags for each season (Figure 9): the GFDL and IPSL families, with lags on the order of 8 days for Region 1, 10 days for Region 2 and up to 40 days for Region 3; the MIROC and MPI families, with no lag in Region 1, a relatively small lag of 15 days in Region 3, and a large lag of 20 days in Region 3; and the CCSM, CMCC and CNRM families with typically smaller lags in all

regions (with the exception of CCSM in Region 1). This suggests that the lag in temperature is not a function of resolution only, but also the result of dynamics inherent to most GCMs.

For the seven CMIP5 GCMs for which ocean mixed layer depth was available, we also compared the magnitude of the lag to the bias in mixed layer depth (Figure 9, right). Our hypothesis was that a deeper mixed layer could moderate seasonal temperature changes, driving the lag in GCM simulations. However, as shown in Figure 9 (right), no such relationship is evident from the 7 GCM simulations available for this analysis. This is not to say that mixed layer depth plays no role at all, but does show that it is not the driving factor for the temperature lag in these models, implying that more attention should be paid to atmospheric drivers of temperature change.

In summary, we found that most CMIP3 and CMIP5 GCMs were able to reasonably simulate the annual cycle of the temperature, but with a systematic delay of approximately one month in the timing of the annual cycle, particularly for the ocean-dominated Regions 2 and 3. In contrast to precipitation, no single model or even a group of models could be identified as more or less able to reproduce the climatology of temperature over the Caribbean than others. The only exceptions are the MRI-CGCM models (the brown bars in Figure 9 left), which incorporate a flux adjustment to bring atmosphere-ocean heat fluxes into balance that the other models do not use.

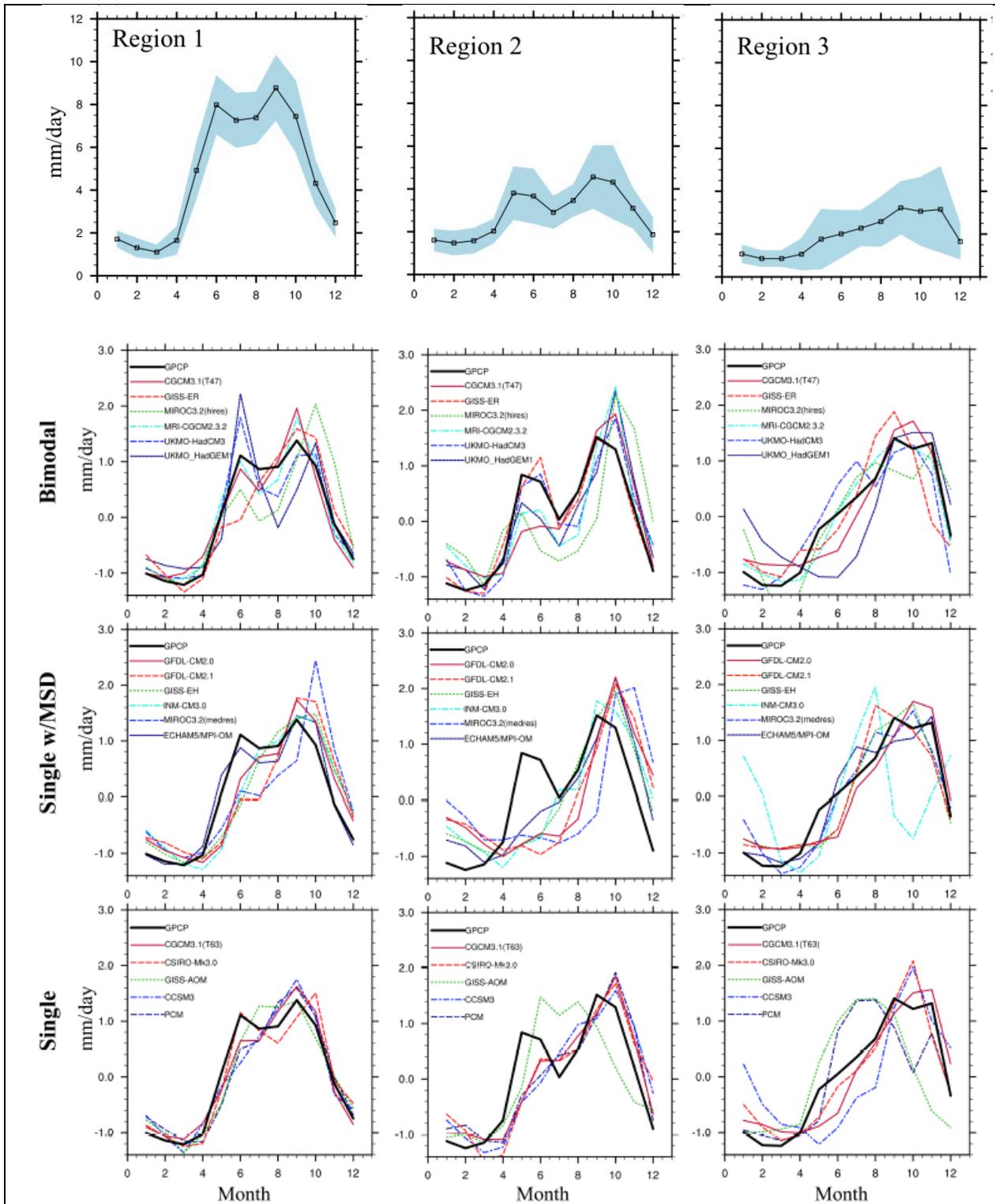
Comparing the climatological annual cycles of observed vs. modeled SST (not shown) reveals that GCM-simulated SSTs also lag observations during late spring and early summer, then remain artificially elevated from the end of summer through the end of the year. In other words, the systematic difference between GCMs simulations of surface air temperature and observations closely mirror the systematic errors of the GCMs in simulating SST as compared to observations, suggesting that model deficiencies in simulating the timing and magnitude of the eastward extension of the warm SST pool over the Caribbean Sea and Gulf of Mexico in summer is likely the result of the same dynamical limitations in GCM simulations as biases in the annual cycle of 2m air temperature over the region.

*A manuscript summarizing the results of this analysis is currently in preparation.*

### **Global Climate Model Performance: Caribbean Precipitation**

**For precipitation, GCMs could be divided into 3 distinct groups based on their ability to simulate the SST and atmospheric circulation features controlling precipitation variability over the region. A greater proportion of CMIP5 GCMs were able to reproduce these patterns as compared to CMIP3, suggesting that the ability of the CMIP ensembles as a whole has improved over time.**

To analyze characteristics of precipitation over the Caribbean, we again divided the Caribbean into the same three regions as above (Figure 6) and conducted a comprehensive literature review and analyzed composite maps of gridded observed and reanalysis-based ocean, surface, and upper-air fields to identify the large-scale atmospheric and oceanic features most relevant to precipitation over the region. We also compared gridded GPCP precipitation to station-based records to quantify differences between broad-scale averages and individual locations before the GPCP dataset was used to evaluate GCM simulations.



**Figure 10.** Observed (top) and GCM-simulated (bottom) annual cycle of precipitation over the three Caribbean regions shown in Figure 6. GCM simulations are divided into the three categories described in the text: models with two precipitation peaks and a mid-summer drought; models with a mid-summer drought and single precipitation peak; and models with one precipitation peak.

The annual cycle of precipitation in the central and eastern Caribbean is characterized by a bimodal structure of two peaks in spring/early summer and fall framing a midsummer drought, while the western Caribbean experiences only the second of these two peaks (Magaña et al. 1999; Jury et al. 2007; Gamble et al. 2008). Previous research has shown temporal variability to be closely related to aspects of the large-scale environment including the Caribbean Low-Level Jet (CLLJ), the North Atlantic Subtropical High (NASH), and SST over various time scales from intra-seasonal to inter-annual (Giannini et al. 1998; Wang 2007; Munoz et al. 2008; Cook and Vizzy 2010; Martin and Schumacher 2010).

Comparing GCM simulations of seasonal precipitation to those observed for each region, we found that not all models are able to simulate the bi-modal precipitation structure seen in Regions 1 and 2 (Figure 10). Based on model performance in Region 2, where the bi-modal structure of annual precipitation is most marked, we divided both CMIP3 and CMIP5 models into three categories:

1. Models that correctly simulate a bimodal distribution with two rainfall peaks in May-June and September-October, punctuated by a mid-summer drought (MSD) in July-August (“good”)
2. Models that reproduce the MSD and the second precipitation maxima only (“fair”)
3. Models that simulate only one precipitation maxima, beginning in early summer (“poor”)

Table 3 lists each of the models in CMIP3 and CMIP5 we evaluate that correspond to the three groups described above. One model (CMIP3 CNRM-CM3) could not be categorized by any of these groups.

For each group, we then examined composite (model average) fields of geopotential height, sea surface temperature, and winds to see if the precipitation-based categories had a common dynamical basis. The answer was yes: there was a noticeable difference between the timing and magnitude of the large-scale dynamics simulated by models in these three different categories, and those dynamical differences were entirely consistent with the observed biases in precipitation. In other words, the biases in dynamics would be expected to lead to exactly the biases in precipitation that were observed in each group of models, and this result was consistent across both CMIP3 and CMIP5 models.

The differences between the model groups are related to the timing and magnitude of the westward extension of the North Atlantic Subtropical High (NASH) and the eastward extension of a warm pool of sea surface temperature (SST) across the Caribbean Sea and the Gulf of Mexico in summer. Specifically, models in category 2 tend to anticipate the westward expansion of the NASH into the Caribbean in early summer. Early onset of NASH results in strong (dry) moisture divergence and drought-like conditions at the time of the May-June observed precipitation peak. Models in category 3 tend to have cooler SST across the region, particularly over the central Caribbean and the Gulf of Mexico, as well as a weaker Caribbean low-level jet accompanying a weaker NASH. In these models, observed June-like patterns of (wet) moisture convergence in the central Caribbean and the Central America and (dry) divergence in the east Caribbean and the Gulf of Mexico persist through September.

**Table 3.** CMIP3 and CMIP5 models corresponding to the three categories of models, identified based on ability to simulate the seasonal cycle of precipitation over the Caribbean Greater Antilles. GCM names correspond to those listed in Table 1. The CNRM-CM3 model was unable to be categorized.

“Good” models able to simulate two precipitation peaks and a mid-summer drought	“Fair” models able to simulate a mid-summer drought and one precipitation peak	“Poor” models able to simulate just one precipitation peak
<b>CMIP3</b>		
CGCM3 T47 (medium resolution)	MIROC 3.2 (medium resolution)	CGCM3 T63 (higher resolution)
MIROC 3.2 (higher resolution)	GFDL CM2.0	CSIRO-Mk3.0
HadCM3	GFDL CM2.1	GISS-AOM
HadGEM1	INMCM 3.0	CCSM3
MRI-CGCM 2.3.2	ECHAM5	PCM
GISS-ER	GISS-EH	
<b>CMIP5</b>		
CanCM4	MIROC-ESM, CHEM	GFDL CM3
MIROC5, 4h	GFDL-ESM2G, M	GISS-E2-H, H-CC, R, R-CC
CNRM-CM5	CCSM4	INMCM4
CSIRO-Mk3.6.0	CESM1 (BGC, WACCM)	
HadCM3		
HadGEM2-AO, CC, ES		
MPI-ESM-LR, MR		
MRI-CGCM3		
CESM1 (CAM5)		

In CMIP3, models were divided evenly between the three groups. In CMIP5, the majority of models showed improvements, moving into a “better” group or (for new models such as CESM1-CAM5) even debuting in Group 1. This suggests that GCM ability to simulate these important large-scale drivers of precipitation over the Caribbean is increasing over time.

A more detailed description of this analysis, discussion of results, and figures showing the differences in NASH, SST and moisture convergence/divergence between the three model groups is provided by Ryu and Hayhoe (2013), attached here as **APPENDIX B**.

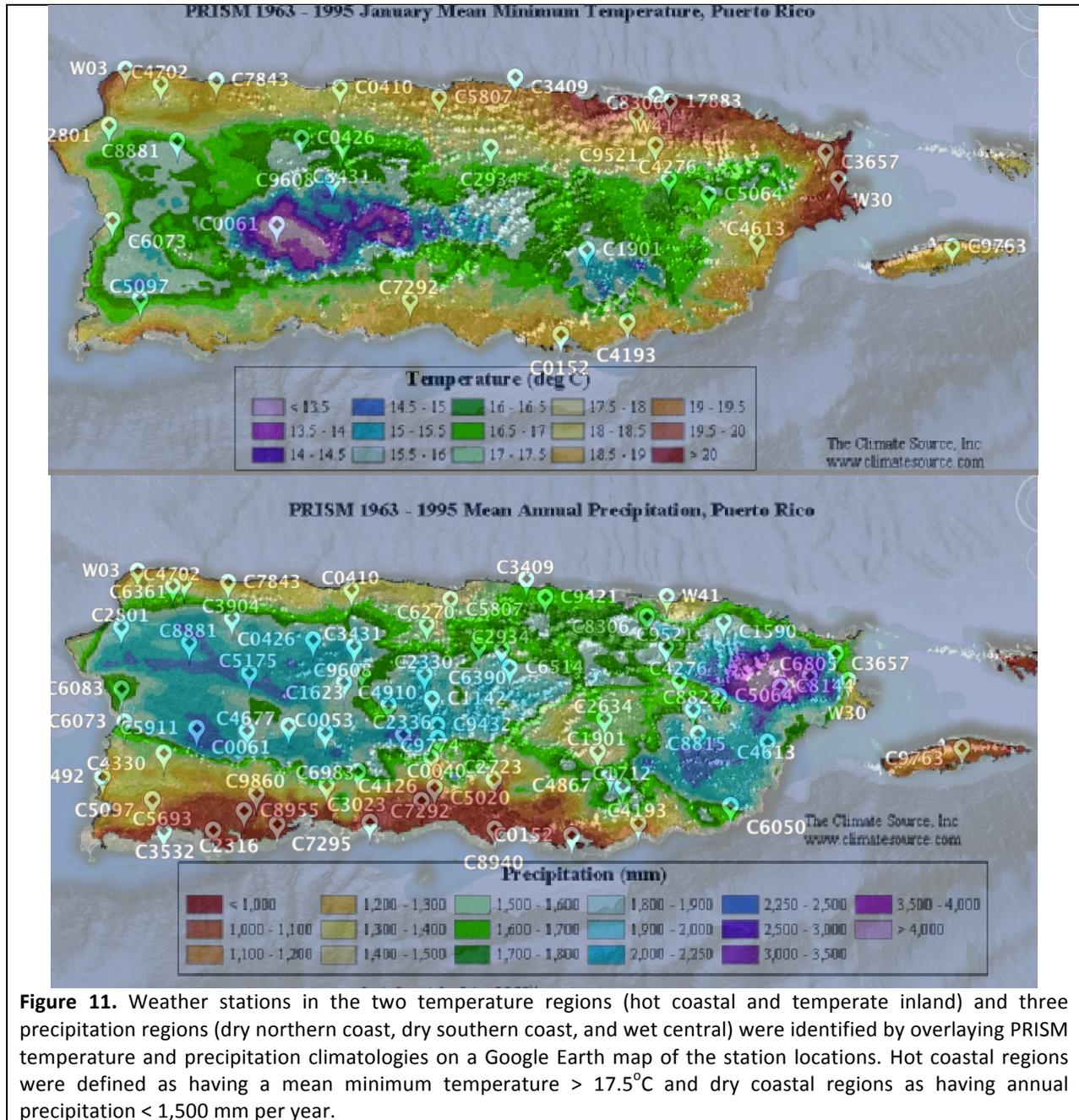
### Climate Projections for Puerto Rico

**Projected changes in 85 temperature- and precipitation-related indicators for Puerto Rico show increases in mean and extreme temperatures, decreases in annual precipitation and moderately-extreme rainfall days, and increases in both dry days and the frequency of extreme precipitation days. Projections from CMIP5 vs. CMIP3 models differ in the magnitude of projected change for some indicators, but agree on the direction of change. Most changes are projected to be greater under higher levels of global mean temperature change.**

Using the station-based high-resolution daily temperature and precipitation projections developed in Step One of this analysis, a broad suite of climate indicators were calculated

for Puerto Rico stations to characterize projected temperature and precipitation changes over the region. To synthesize these results, they were averaged three ways:

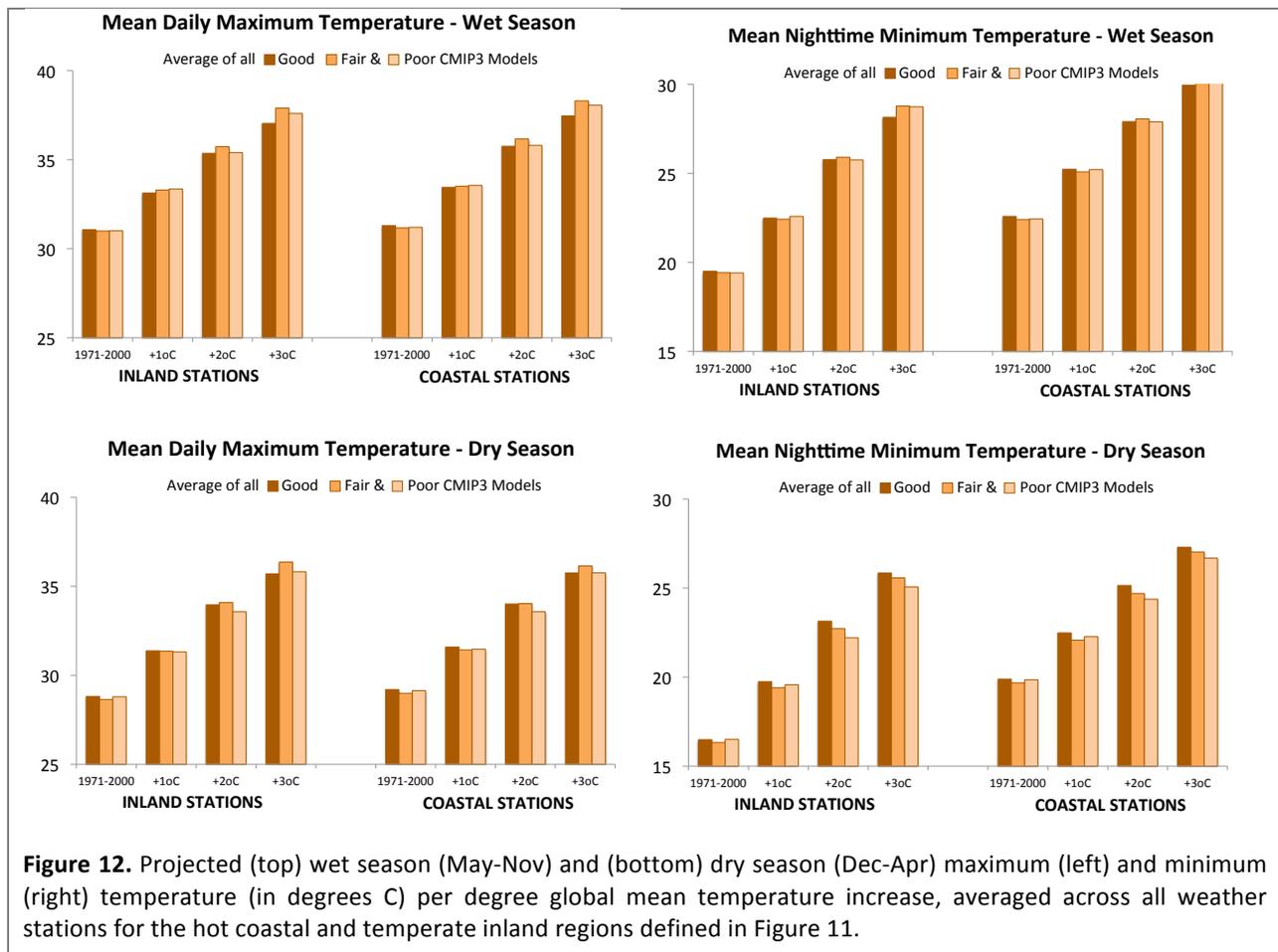
1. Over regions, with the island of Puerto Rico being divided into two regions, hot coastal and temperate inland, for temperature and three regions, dry northern coast, dry southern coast, and wet central region, for precipitation (Figure 11)
2. Over GCMs, with the model categories listed in Table 3 being used to determine which model simulations could be combined
3. Over global mean temperature change, for increases of +1, 2 and 3°C relative to the 1971-2000 global mean temperature simulated by each GCM/scenario combination.



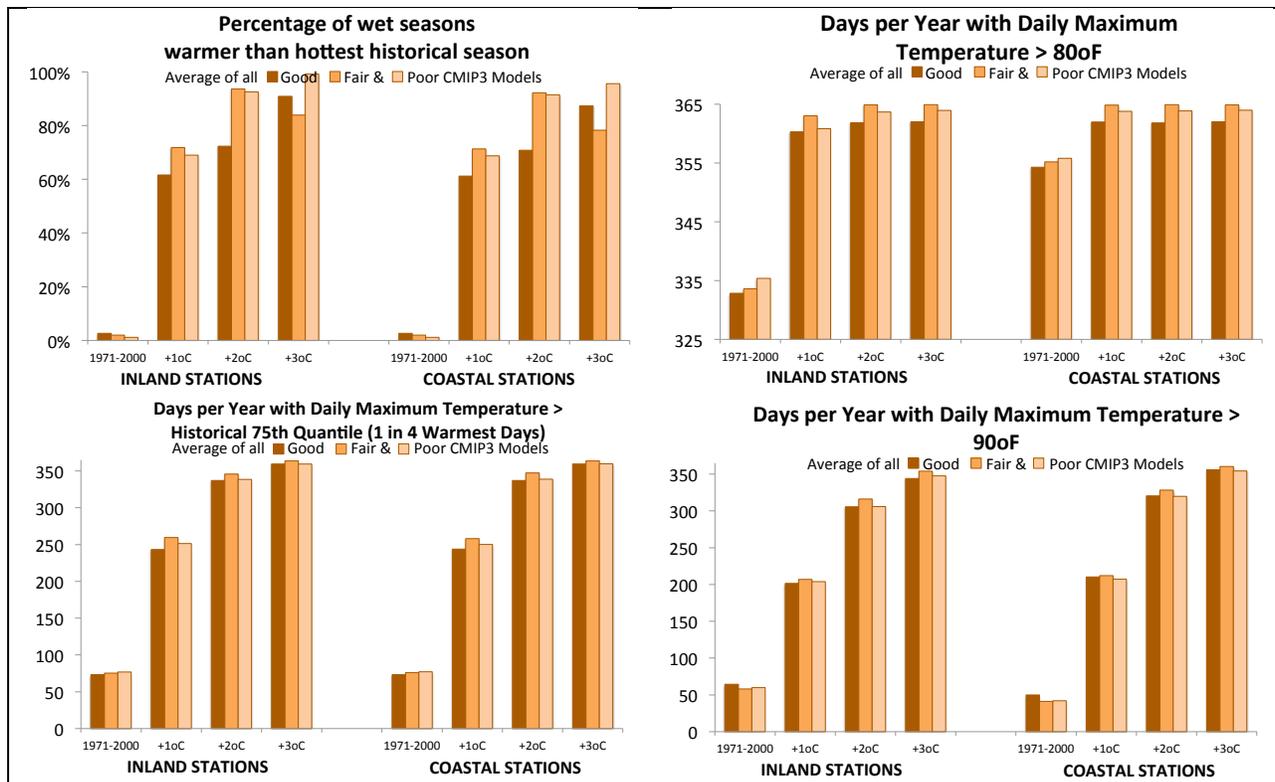
**Figure 11.** Weather stations in the two temperature regions (hot coastal and temperate inland) and three precipitation regions (dry northern coast, dry southern coast, and wet central) were identified by overlaying PRISM temperature and precipitation climatologies on a Google Earth map of the station locations. Hot coastal regions were defined as having a mean minimum temperature > 17.5°C and dry coastal regions as having annual precipitation < 1,500 mm per year.

As global mean temperature increases, temperatures in Puerto Rico are also projected to increase. Projected changes are slightly higher for the dry season as compared to the wet season, and for minimum nighttime temperature as compared to maximum daytime temperature (Figure 12). Consistent with the results reported above, that the GCMs cannot be differentiated based on their ability to simulate temperature, there does not appear to be a significant difference between good vs. fair or poor GCMs for most indicators. There is a small difference between projected temperatures in the wet season, with “good” GCMs producing a slightly smaller temperature increase than “fair” or “poor” models.

Projected temperature increases are greater than the global average, with a simulated ~2-3°C increase in local temperature per 1°C increase in global mean temperature. In terms of temperature variability, the standard deviation and the seasonal range in both daytime and nighttime temperature is projected to increase, particularly for coastal stations as compared to inland and in the wet season as compared to the dry season. Temperature extremes are projected to increase at a similar rate to mean temperatures, with the daily maximum temperature on the hottest days of the year estimated to rise by 2-2.5°C per degree global mean temperature change for inland stations and slightly more for coastal locations. In contrast, increases in the coldest nighttime minimum temperatures of the year are projected to be greater for inland stations as compared to coastal. These changes are illustrated in the full set of figures available in **APPENDIX C**.



**Figure 12.** Projected (top) wet season (May-Nov) and (bottom) dry season (Dec-Apr) maximum (left) and minimum (right) temperature (in degrees C) per degree global mean temperature increase, averaged across all weather stations for the hot coastal and temperate inland regions defined in Figure 11.



**Figure 13.** Projected changes in most mean and extreme indicators occur incrementally per degree global mean temperature change, indicating a shift in the mean of the distribution rather than the shape. For a few indicators of “moderate extremes”, however, there is a large increase in projected values under only a +1°C global temperature change. These include the percentage of warm wet seasons (top left), the days per year with maximum temperature exceeding 80 and 90°F (right) and days per year with maximum temperature above the 75<sup>th</sup> quantile or 1-in-4 historical hottest days.

In terms of hot temperatures, perhaps most striking is the fact that, with only a +1°C increase in global mean temperature, more than 50% of wet seasons are expected to be warmer than the warmest historical wet season (Figure 13, top left). Days per year over 80°F at inland stations is expected to see a similar abrupt increase, from approximately 330 to 360 days per year (Figure 13, top right) and days over 90°F increase by approximately 100 days per year for both inland and coastal stations (Figure 13, middle right). A similar abrupt change is seen for the number of days per year greater than the historical 50<sup>th</sup> quantile, or median, and 75<sup>th</sup> quantile, or 1-in-4 hottest days (Figure 13, middle left).

Significant but much more incremental increases are projected for days over 95 and 100°F and for nights below 65 and 75°F and above 80, 85 and 90°F. Similarly, for days per year over the 90<sup>th</sup>, 99<sup>th</sup> and 99.9<sup>th</sup> quantile (i.e., exceeding the historical 1-in-10, 1-in-100 and 1-in-1000 hottest days) as well as for most projected changes in cold days, increases (decreases) are projected to occur evenly per degree global mean temperature increase (see **APPENDIX C**).

Cooling degree-days are often taken as a measure of energy demand for air conditioning. A threshold is established at which the air conditioning is turned on (here, two thresholds are used: 75°F and 85°F) and the cooling load calculated by estimating the number of hours

and degrees above those temperatures that occur during the course of a year. Historically, annual cooling degree-days average around 400 for temperate inland locations and 800 for warmer coastal locations, using a 75°F threshold. With a +1°C increase in global mean temperature, inland locations are projected to experience an average of more than 1000 degree-days per year; coastal locations, more than 1500. The difference between the two regions decreases as temperature increases; under a +3°C global mean temperature change, inland locations are expected to experience an average of 3,000 degree-days per year while coastal locations would expect an average of 3,200 degree-days per year (see **APPENDIX C**).

For a water-limited region such as the Caribbean, climate change is not just about warmer temperatures; as the planet warms, precipitation patterns are also expected to shift in both space and time. Changes in precipitation could have an equal or even greater impact on the local ecosystems, economy, and water resources as increases in temperature. Changes in the intensity and frequency of heavy rainfalls, as well as the duration of dry periods, would also affect agriculture and water supply as well as risk of flood and drought.

Precipitation in Puerto Rico historically averages around 900 to 1000 mm in the wet season (May-Nov) and 450 mm in the dry season (Dec-Apr), with a sharp distinction between total rainfall in drier coastal areas vs. wetter inland regions (see Figure 11).

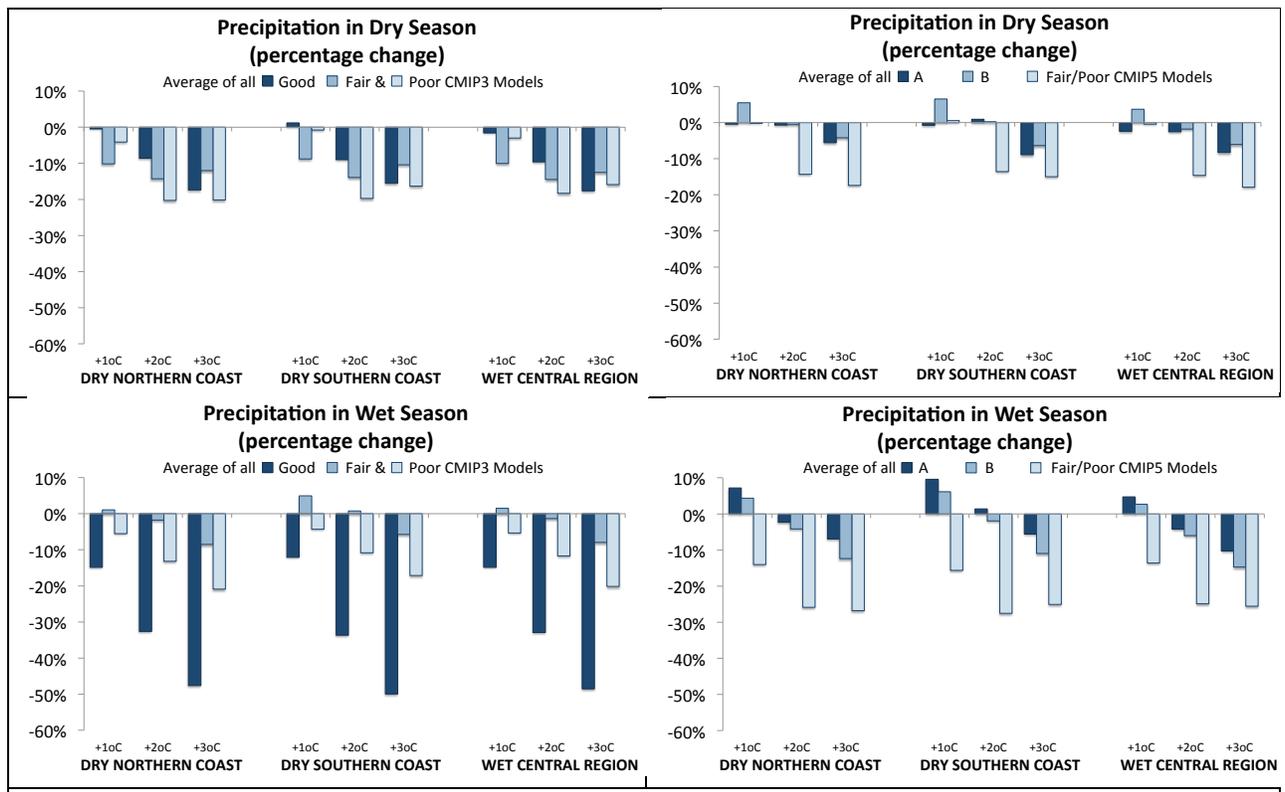
For the dry season under future climate change, both CMIP3 and CMIP5 simulations consistently project little change for a +1°C increase in global mean temperature but decreases in dry season precipitation on the order of -5% for a +2°C increase in global mean temperature and -10 to -15% for a +3°C increase (Figure 14, top). Projected changes from “good” GCMs are generally similar to or smaller than projected changes from GCMs in the “fair” and smaller than projected changes from GCMs in the “poor” categories, suggesting that model performance does play a role in moderating the magnitude, but not the sign, of future change in dry season precipitation.

Projected changes in wet season precipitation are more challenging to interpret (Figure 14, bottom). For a +1°C change, CMIP5 “good” models project a small increase in wet season precipitation on the order of 5-10%, consistent with recent observed changes in the historical record (NCA, 2012). In contrast, CMIP3 models project a decrease on the order of 10-15% under the same amount of global change. For global mean temperature changes of +2 and +3°C, both CMIP3 and CMIP5 “good” models project decreases in precipitation, but for CMIP3 decreases from “good” models are *greater* than from other GCMs (-30% for +2°C and -50% for +3°C from “good” GCMs; projected changes from CMIP3 “fair” and “poor” models are inconsistent), while for CMIP5 projected decreases from “good” models are *smaller* than those projected by the combined projections of “fair” and “poor” models: on the order of -5% for +2°C and -10% for +3°C.

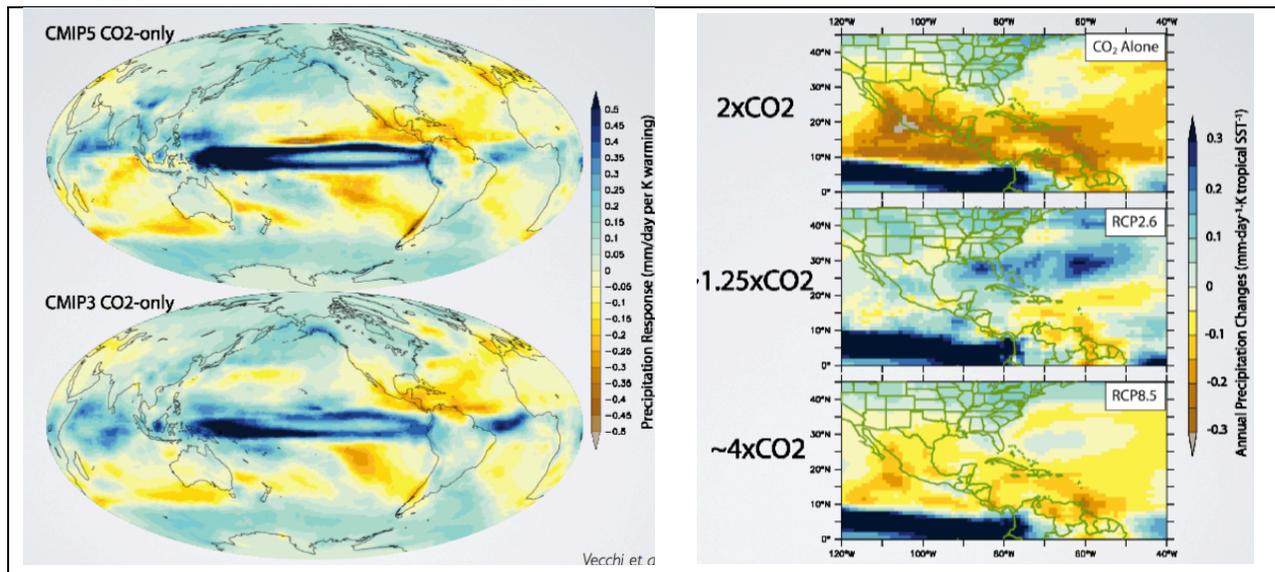
These results raise a crucial question: why is the magnitude of projected changes from “good” CMIP3 vs. “good” CMIP5 models so different, when projected changes in dry season precipitation are similar and model ability to simulate the large-scale drivers of precipitation variability is also similar (see **APPENDIX B**)? Fortunately, research by Gabe Vecchi and colleagues from the Geophysical Fluid Dynamics Laboratory suggests a plausible reason. When future simulations by CMIP3 and CMIP5 models are driven by

changes in carbon dioxide alone, the spatial pattern and magnitude of resulting changes in precipitation are quite similar, including projected decreases in precipitation over Puerto Rico (Figure 15, left). However, when CMIP5 models are driven by the new RCP scenarios (as compared to CMIP3 simulations driven by the SRES scenarios), projected rainfall changes are substantially lower over Puerto Rico (-0.1 to -0.15 mm/day) even under an RCP scenario with a 4x increase in carbon dioxide levels as compared to a projected increases under a doubling of carbon dioxide alone (-0.25 to -0.3 mm/day).

This implies that it is not the CMIP3 vs. CMIP5 model response to carbon dioxide that is so different, but rather the aerosol inputs and model response to those inputs. Vecchi et al. (2012) specifically hypothesize that the differences in precipitation changes between the SRES and carbon dioxide-only scenarios, which are similar, and the RCP scenarios, which are very different, could be the result of very different projected changes in aerosols in the RCP vs. SRES scenarios, compounded by the fact that many CMIP5 models now contain more sophisticated treatments of aerosol chemistry and their impact on clouds than did previous CMIP3 GCMs; so a change in aerosol emissions in the input scenarios will have a proportionally greater effect now than they would with the previous generation of models. While this “aerosol effect” has global implications, Vecchi et al. highlight the Caribbean as being one of the regions most affected in terms of altering projected precipitation patterns (Figure 15, right).



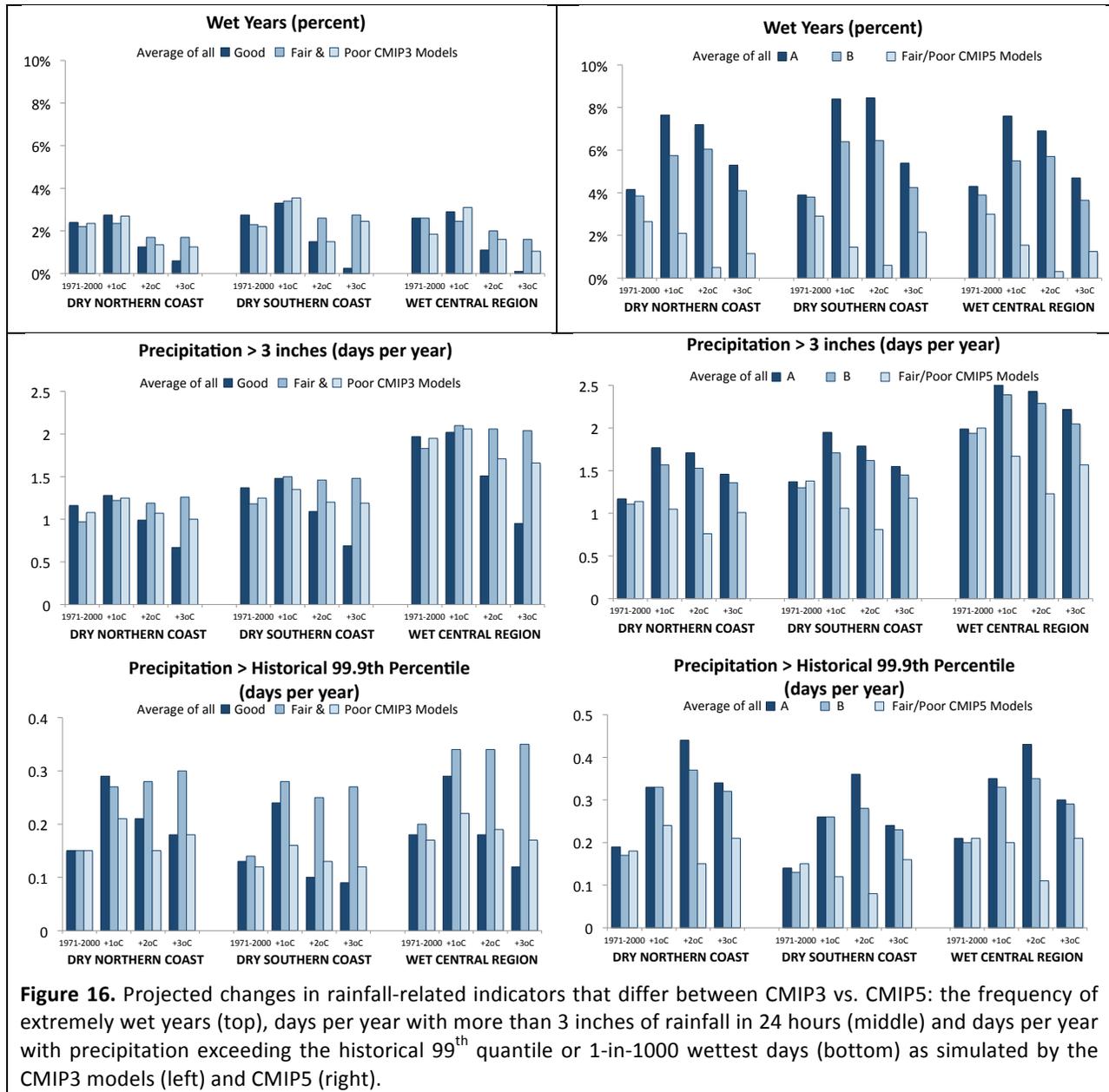
**Figure 14.** Projected changes dry season (top) and wet season (bottom) precipitation as simulated by CMIP3 models (left) and CMIP5 models (right). For CMIP5 plots, projected changes the “fair” and “poor” categories were combined as each had only 2 models. Instead, projected changes in “good” models are plotted for the full CMIP5 ensemble (A) and for a sub-set of the CMIP5 “good” GCMs with matching models in the CMIP3 ensemble (B).



**Figure 15.** Left: projected changes in global precipitation per degree global mean temperature change, as simulated by CMIP5 models (top) and CMIP3 models (bottom) driven by increases in carbon dioxide alone. Right: projected changes in precipitation over the Caribbean as simulated by CMIP5 models driven by carbon dioxide alone (top), by low carbon dioxide and low aerosol emissions (middle), and by high carbon dioxide and medium aerosol emissions (bottom). Figures from Vecchi, 2012 available online at: [http://www.gfdl.noaa.gov/cms-filessystem-action?file=user\\_files/gav/presentations/2012/vecchi\\_cmip5\\_princeton.pdf](http://www.gfdl.noaa.gov/cms-filessystem-action?file=user_files/gav/presentations/2012/vecchi_cmip5_princeton.pdf)

Differences between CMIP3/SRES and CMIP5/RCP-based projections of wet season precipitation also lead to differences in certain precipitation-related indicators, while for other indicators the results are similar across both sets of model simulations. One key difference is the projected frequency of wet years. Initially, little change in the frequency of wet years is expected for CMIP3 simulations, then a decrease by +3°C global mean temperature change (Figure 16, top left). In contrast, for CMIP5 simulations there is a near-doubling of wet years followed by a slow decline (Figure 16, top right). Another difference is in projected changes in extremely heavy precipitation days, such as more than 2 or 3 inches per day, or the number of days exceeding the historical 1-in-100 or 1-in-1000 wettest days. CMIP3 “good” models project a small increase followed by a sharp decrease, while CMIP5 “good” models project a large increase followed by a slow decrease (Figure 16, middle and bottom rows).

In terms of similarities, both CMIP3 and CMIP5 show a slow decline in the standard deviation of dry-season rainfall and little change in wet-season rainfall (see **APPENDIX C**). Neither suggest much change in the average intensity of precipitation during either the wet or the dry season (CMIP3 “good” models do project a slight decrease in the wet season, likely driven by the large decrease in projected overall rainfall during that season). Both sets of models show a small increase in the amount of precipitation falling on the wettest days of the year, followed by a small decrease that is more gradual for CMIP5 than for CMIP3. The models also agree on a slight increase in the number of dry days during the dry



season and a larger increase in dry days during the wet season, as well as moderate increases, on the order of a few days per year, in the length of the longest and the average dry periods in each season and in the total number of wet days per year. There is also good agreement on projected decreases in “moderately extreme” precipitation such as the number of days per year with more than 1 inch of precipitation, or days per year with precipitation greater than the historical 1-in-4 wettest days (see **APPENDIX C**).

What can we conclude? Clearly, model performance, in terms of ability to simulate large-scale drivers of precipitation variability, does matter. For both CMIP3 and CMIP5, projected changes from “good” models are noticeably different than changes projected by models in other groups (see Figures 14 and 16; also **APPENDIX C**). However, *for wet season rainfall*

*over the Caribbean*, the benefit of selecting better-performing models appears to be currently outweighed by the benefit of selecting models and input scenarios that do (CMIP5 and RCP) vs. those that do not (CMIP3 and SRES) take into account projected changes in aerosol emissions and their impact on cloud formation and precipitation.

For some variables--including precipitation for some regions, such as South America or Australia (Figure 15), and even for some seasons, such as the dry season over Puerto Rico or projected changes in dry days or “moderately extreme” precipitation events such as days with >1 inch of rain or rainfall exceeding the historical 1-in-4 wettest day--CMIP3 and CMIP5 models may be used as a sort of “super-ensemble” that yield approximately equivalent results. However, **for wet season rainfall for Puerto Rico and the Caribbean and for extreme precipitation greater than 1 inch in 24 hours, these results strongly suggest that CMIP5 projections from “good” models able to simulate the dynamics driving the Caribbean wet season and the mid-summer drought should be used in place of CMIP3 projections, even though “good” CMIP3 models are able to reproduce observed precipitation-related dynamics over the Caribbean.** It is likely that CMIP3 results over-estimate projected decreases in wet season rainfall and in extreme precipitation events because they are driven primarily by carbon dioxide rather than a combination of carbon dioxide and aerosols.

## CONCLUSIONS AND RECOMMENDATIONS

This project evaluated global climate model simulations of temperature and precipitation over the Caribbean, generated a dataset of high-resolution climate change projections for individual weather stations throughout the region, and analyzed projected changes for the island of Puerto Rico.

For annual changes in **temperature**, nearly every CMIP3 and CMIP5 GCM shows a lag of days to weeks compared to observations, manifested by a consistent under-estimation of Northern Hemisphere late spring/summer temperatures and over-estimation of fall/winter temperatures. Comparison of observed with model-simulated SST suggests that the systematic temperature differences are closely related to model ability to simulate the timing of the northeastward expansion of a warm SST pool from the Pacific coast of Central America up to Cuba in June, throughout the Gulf of Mexico in summer, and over to the edge of the West Indies by fall. This causes most GCMs to consistently under-estimate late spring/summer SST and over-estimate fall/winter SST in the Central Caribbean, the same bias seen in 2 meter temperature.

*The recommended course of action for future temperature projections is to use as many global climate models as is feasible for any given impact analysis, and not to attempt to select a sub-set of “better performing” models.*

For **precipitation**, there was a noticeable difference in the ability of GCMs to simulate the bi-modal structure of precipitation in the central and eastern Caribbean, which is characterized by two peaks in early summer and fall, punctuated by a mid-summer drought. When the large-scale atmospheric and oceanic conditions were compared between different model groups, the reason why some GCMs failed to produce observed shifts in precipitation was evident: one group of models shows the NASH moving over the region too early, suppressing the first precipitation peak, while another group shows it moving over too late and too weakly, unable to produce a mid-summer drought. CMIP3 and CMIP5 models can both be grouped into the same categories, although for CMIP3 approximately 1/3 of the models fall into each of the three categories (good, fair, and poor) while the majority of CMIP5 models are characterized as “good”. Future rainfall projections show significant differences in terms of the magnitude (although not the direction) of change predicted for seasonal and extreme precipitation, depending on which category the model falls under.

*The recommended course of action is to use global climate models that have been proven capable of simulating the processes that drive changes in precipitation over the Caribbean. A list of these models is provided in Table 3.*

For **precipitation**, an additional complication was identified when comparing CMIP3 vs. CMIP5-simulated projections of wet season precipitation. Specifically, we found a larger difference between seasonal and extreme precipitation projections simulated by “good” CMIP3 models as compared to “good” CMIP5 models than between the projections simulated by “good” models vs. “poor” models within a given modeling group. It is likely (although not yet proven) that these differences are due to CMIP3-based projections primarily being driven by increases in carbon dioxide, while CMIP5-based projections are

driven by changes in both carbon dioxide and aerosols, which can affect cloud properties and hence rainfall.

*Because of the substantial differences in the projected changes in wet season and extreme rainfall that result from CMIP3 vs. CMIP5 simulations, the recommended course of action is to preferentially use CMIP5 simulations alone for any projections of changes in wet season rainfall or extreme rainfall.*

**Because of this conclusion, the scope of the original project was substantially expanded to include generation and provision of projections of daily rainfall for the period 1950-2100 as simulated by CMIP5 models.**

**Future projections** for long-term weather stations for Puerto Rico show decreases in both wet- and dry-season precipitation, slight decreases in days with moderately extreme precipitation (e.g. rainfall > 1 inch in 24 hours, or rainfall > the 1-in-4 historical wettest day), and increases in days with more extreme rainfall (e.g. rainfall > 3 inches in 24 hours, or rainfall > the 1-in-100 historical wettest day) as well as in the number of dry days per year and the duration of multi-day dry periods. Temperatures are projected to increase, at a similar rate for both mean or average temperatures and extremes.

*Climate model simulations agree that Puerto Rico can expect temperature increases coupled with moderate decreases in average rainfall and increases in the frequency of extreme rainfall events.*

## OUTREACH

The methods used in this project and the results of the climate model inter-comparison and evaluation have been incorporated into presentations made to a broad range of professional audiences, from non-profit organizations such as the National Wildlife Federation to federal agencies such as the Fish and Wildlife Service and academic collaborators in the Climate Science Center network.

The projections generated by this project are expected to be broadly used throughout the Caribbean region.

This research has resulted in one article in press (see **APPENDIX B**: Ryu and Hayhoe, 2013) and three articles in preparation, describing: (1) the evaluation of GCM ability to simulate variability in temperature over the Caribbean; (2) climate projections for Puerto Rico and the Caribbean; and (3) comparing the importance of model performance vs. model generation in quantifying uncertainty in future rainfall changes over the Caribbean.

## REFERENCES

- Adler R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin. (2003) The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology* 4, 1147-1167.
- Cook, K. & E. Vizi. (2010) Hydrodynamics of the Caribbean Low-Level Jet and Its Relationship to Precipitation. *Journal of Climate* 23, 1477-1494.
- Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor (2003) An overview of results from the Coupled Model Intercomparison Project, *Global and Planetary Change*, 37(1-2) 103-133.
- Dettinger, M. D., D. R. Cayan, M. K. Meyer, and A. E. Jeton (2004) Simulated Hydrologic Responses to Climate Variations and Change in the Merced, Carson, and American River Basins, Sierra Nevada, California, 1900–2099, *Climatic Change*, 62(1-3) 283-317.
- Enfield, D. & E. J. Alfaro (1999) The Dependence of Caribbean Rainfall on the Interaction of the Tropical Atlantic and Pacific Oceans. *Journal of Climate* 12, 2093–2103.
- Gamble, D., D. Parnell & S. Curtis (2008) Spatial variability of the Caribbean mid-summer drought and relation to north Atlantic high circulation. *International Journal of Climatology* 27 28, 343-350.
- Giannini A, Y. Kushnir & M. Cane. (2000) Interannual variability of Caribbean Rainfall, ENSO, and the Atlantic Ocean. *Journal of Climate* 13, 297-311.
- Hay, L. E. (2000) A comparison of delta change and downscaled GCM scenarios for three, *Journal of the American Water Resources Association* 36(2), 387-397.
- Hayhoe, K., C. Wake, B. Anderson, X.-Z. Liang, E. Maurer, J. Zhu, J. Bradbury, A. DeGaetano, A. Hertel, and D. Wuebbles (2008) Regional Climate Change Projections for the Northeast U.S., *Mitigation and Adaptation Strategies for Global Change* 13, 425-436.
- Hayhoe, K., et al. (2004) Emissions pathways, climate change, and impacts on California, *Proceedings of the National Academy of Sciences of the United States of America*, 101(34) 12422-12427.
- Hayhoe, K., K. Dixon, A. Stoner, J. Lanzante and A. Radhakrishnan. (2012) Is the past a guide to the future? Evaluating the assumption of climate stationarity in statistical downscaling. *Presented at the 2012 Fall Meeting of the American Geophysical Union*.
- Jury M., B. Malmgren & A. Winter. (2007) Subregional precipitation climate of the Caribbean and relationships with ENSO and NAO. *Journal of Geophysical Research* 112, D16107, doi:10.1029/2006JD007541.
- Jury, M. (2009) A quasi-decadal cycle in Caribbean climate. *Journal of Geophysical Research* 114, D13102.
- Knutti, R. (2010) The end of model democracy? *Climatic Change* 102, 395–404, doi:10.1007/s10584-010-9800-2
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak and G. Meehl. (2010) Challenges in combining projections from multiple climate models. *Journal of Climate* 23, 2739-2758
- Koenker R, Bassett G Jr. (1978) Regression quantiles. *Econometrica* 46, 33–50.

- Magaña V., J. Amador & S. Medina. (1999) The midsummer drought over Mexico and Central America. *Journal of Climate* 12, 1577-1588.
- Malmgren, B., A. Winter & D. Chen (1998) El Niño–Southern Oscillation and North Atlantic Oscillation Control of Climate in Puerto Rico. *Journal of Climate* 11, 2713–2717.
- Martin, E. & C. Schumacher. (2011) The Caribbean Low-Level Jet and its Relationship with Precipitation in IPCC AR4 Models. *Journal of Climate* 24, 5935-5950.
- Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer & K. E. Taylor. (2007) The WCRP CMIP3 multi-model dataset: A new era in climate change research, *Bulletin of the American Meteorological Society* 88, 1383-1394.
- Monterey, G. & S. Levitus. (1997) Climatological cycle of mixed layer depth in the world ocean. *U.S. Gov. Printing Office, NOAA NESDIS*, 5pp. Available online at: <http://www.esrl.noaa.gov/psd/data/gridded/data.nodc.woa94.html>
- Moss, R., J. Edmonds, K. Hibbard, M. Manning, S. Rose, D. van Vuuren, T. Carter, S. Emori, M. Kainuma, T. Kram, G. Meehl, J. Mitchell, N. Nakicenovic, K. Riahi, S. Smith, R. Stouffer, A. Thomson, J. Weyant & T. Wilbanks. (2010) The next generation of scenarios for climate change research and assessment. *Nature* 463, 747-756 (11 February 2010)
- Muñoz E., A. Busalacchi, S. Nigam & A. Ruiz-Barradas. (2008) Winter and summer structure of the Caribbean low-level jet. *Journal of Climate* 21, 1260-1276.
- Nakicenovic, N. et al (2000). *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, U.K., 599 pp.
- National Research Council Committee on Stabilization Targets for Atmospheric Greenhouse Gas Concentrations (2010) *Climate Stabilization Targets: Emissions, Concentrations, and Impacts over Decades to Millennia*. Cambridge University Press, 190 pp.
- Randall, D., R. Wood, S. Bony, R. Colman, T. Fichfet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. Stouffer, A. Sumi and K. Taylor (2007) *Climate Models and Their Evaluation*. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Rauscher, S., F. Giorgi, N. Diffenbaugh & A. Seth. (2008) Extension and Intensification of the Meso-American mid-summer drought in the twenty-first century. *Climate Dynamics* doi 10.1007/s00382-007-0357-1.
- Reynolds, R., N. Rayner, T. Smith, D. Stokes & W. Wang. (2002) An improved in situ and satellite SST analysis for climate. *Journal of Climate* 15, 1609-1625.
- Stoner, A., K. Hayhoe, X. Yang and D. Wuebbles (2012) An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology* doi: 10.1002/joc.3603
- Taylor, K., R. Stouffer & G. Meehl. (2012) An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society* 93, 485–498.
- U.S. Global Change Research Program (USGCRP, 2000) *US National Assessment of the Potential Consequences of Climate Variability and Change*. Cambridge University Press.

- U.S. Global Change Research Program (USGCRP, 2009) *Global Climate Change Impacts in the United States*. Cambridge University Press. Available online at: <http://www.globalchange.gov/usimpacts/>.
- UK Meteorological Office. (2013) *Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)*, [Internet]. NCAS British Atmospheric Data Centre, 2012. Available from [http://badc.nerc.ac.uk/view/badc.nerc.ac.uk\\_ATOM\\_dataent\\_ukmo-midas](http://badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_ukmo-midas)
- Uppala, S. et al. (2005) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* 131, 2961-3012.
- Vecchi, G. (2012) Precipitation Response in Climate Models: What's new in CMIP5? (thus far) *Manuscript in preparation; presentation available at:* [http://www.gfdl.noaa.gov/cms-filesystem-action?file=user\\_files/gav/presentations/2012/vecchi\\_cmip5\\_princeton.pdf](http://www.gfdl.noaa.gov/cms-filesystem-action?file=user_files/gav/presentations/2012/vecchi_cmip5_princeton.pdf)
- Vrac, M., M. Stein, and K. Hayhoe (2007) Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Climate Research* 34(3) 169-184.
- Wang, C. (2007) Variability of the Caribbean Low-Level Jet and its relations to climate. *Climate Dynamics* 29, 4, 411-422.
- Whyte, F., M. Taylor, T. Stephenson & J. Campbell. (2008) Features of the Caribbean low level jet. *Journal of Climatology* 28, 1, 119-128.
- Wood, A. W., R. Leung, V. Sridhar and D. Lettenmaier (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* 62(1-3) 189-216.

## **APPENDIX A. Weather Stations**

This Appendix lists the ID, latitude, and longitude of the weather stations for which sufficient data was available to downscale daily maximum and minimum temperature and precipitation. There are a total of 1,542 (maximum temperature), 1,508 (minimum temperature) and 2,129 (precipitation) individual long-term weather stations in the Caribbean.

*Due to its length, this Appendix is provided as a separate Excel file.*

## **APPENDIX B. Analysis of GCM ability to simulate Caribbean precipitation drivers**

Ryu, J.H. and K. Hayhoe (2013) Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations. *Climate Dynamics* (in press)

# Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations

Jung-Hee Ryu<sup>a</sup> and Katharine Hayhoe<sup>a,b</sup>

*Climate Science Center, Texas Tech University, Lubbock, TX<sup>a</sup>*

*Department of Political Science, Texas Tech University, Lubbock, TX<sup>b</sup>*

*Corresponding author:* Jung-Hee Ryu

Climate Science Center, Texas Tech University

MS1015, Lubbock, TX 79409

Phone: 806-787-0894

[jung-hee.ryu@ttu.edu](mailto:jung-hee.ryu@ttu.edu)

Submitted to

*Climate Dynamics*

February 2013

# Abstract

1  
2 We assess the ability of Global Climate Models participating in phases 3 and 5 of the  
3 Coupled Model Intercomparison Project (CMIP3 and CMIP5) to simulate observed annual  
4 precipitation cycles over the Caribbean. Compared to weather station records and gridded  
5 observations, we find that both CMIP3 and CMIP5 models can be grouped into three  
6 categories: (1) models that correctly simulate a bimodal distribution with two rainfall  
7 maxima in May-June and September-October, punctuated by a mid-summer drought (MSD)  
8 in July-August; (2) models that reproduce the MSD and the second precipitation maxima  
9 only; and (3) models that simulate only one precipitation maxima, beginning in early  
10 summer. These categories appear related to model simulation of the North Atlantic  
11 Subtropical High (NASH) and sea surface temperature (SST) in the Caribbean Sea and Gulf  
12 of Mexico. Specifically, models in category 2 tend to anticipate the westward expansion of  
13 the NASH into the Caribbean in early summer. Early onset of NASH results in strong  
14 moisture divergence and MSD-like conditions at the time of the May-June observed  
15 precipitation maxima. Models in category 3 tend to have cooler SST across the region,  
16 particularly over the central Caribbean and the Gulf of Mexico, as well as a weaker  
17 Caribbean low-level jet accompanying a weaker NASH. In these models, observed June-like  
18 patterns of moisture convergence in the central Caribbean and the Central America and  
19 divergence in the east Caribbean and the Gulf of Mexico persist through September. This  
20 analysis suggests systematic biases in model structure may be responsible for biases in  
21 observed precipitation variability over the Caribbean and more confidence may be placed in  
22 the precipitation simulated by the GCMs that are able to correctly simulate seasonal cycles  
23 of SST and NASH.

24 *Keywords: Caribbean precipitation, Mid-summer drought, North Atlantic Subtropical*  
25 *High, Caribbean low-level jet, sea-surface temperature, climate models*

# 1 **1 Introduction**

2           Surface climate conditions in the Caribbean have been shown to be sensitive to  
3 large-scale ocean and atmosphere dynamics including sea surface temperature, the El Niño-  
4 Southern Oscillation (ENSO), and the Caribbean Low-Level Jet (CLLJ) (e.g., Malmgren et  
5 al., 1998; Enfield et al., 1999; Wang, 2007; Small et al. 2007; Whyte et al., 2008; Jury,  
6 2009a). The spatial distribution of Caribbean precipitation strongly reflects the more  
7 localized effects of island orography. However, temporal variability is closely related to  
8 aspects of the large-scale dynamical features including the CLLJ, NASH, and SST over  
9 various time scales from intraseasonal to interannual (Giannini et al. 2000; Wang 2007;  
10 Munoz et al. 2008; Cook and Vizy 2010). The NASH is the semi permanent subtropical  
11 high pressure system dominating the Atlantic basin in the lower troposphere (Davis et al.  
12 1997). The CLLJ is an easterly jet in the lower troposphere located over the Caribbean Sea  
13 between the northern coast of South America and the Greater Antilles that has a semi-annual  
14 cycle, with maxima in February and July (Stensrud 1996; Amador 1998; Munoz et al. 2008;  
15 Whyte et al. 2008). It is thought that the peak in July is related to the seasonal cycle of the  
16 NASH, while the maximum in February is likely caused by heating over northern South  
17 America (Cook and Vizy 2010).

18           A key characteristic of precipitation in the Caribbean is the annual bimodal  
19 distribution where the summer rainy season is interrupted by a mid-summer drought, or  
20 MSD (Magaña et al. 1999; Jury et al. 2007; Gamble et al. 2008). Current understanding of  
21 the regional-scale atmospheric processes involved in driving the precipitation variability has  
22 been summarized by Gamble and Curtis (2008). A number of studies have examined the  
23 role of atmospheric and oceanic dynamics in driving this feature; it is important to review  
24 these, as we propose to examine the extent to which GCMs are able to simulate large-scale  
25 features related to precipitation and the MSD. One hypothesis is that an enhanced mid-

1 summer CLLJ could be responsible for the MSD because the observed divergence over the  
2 eastern Caribbean (Wang and Enfield 2003; Wang 2007), SST cooling due to strong wind  
3 stress (Giannini et al. 2000), and increased vertical shear (Inoue et al. 2002; Wang and Lee  
4 2007) due to a strengthening of the CLLJ all work together to inhibit convective  
5 development and tropical cyclogenesis. Another complementary hypothesis is that the  
6 CLLJ strengthens in response to a westward shift in the NASH; the CLLJ itself is basically  
7 geostrophic and its seasonal cycle is associated with the meridional geopotential height  
8 gradient (Cook and Vizy 2010). Previous studies examining the relevance of the NASH to  
9 the midsummer drought over the Caribbean (Curtis and Gamble 2007; Small et al. 2007;  
10 Munoz et al. 2008; Gamble et al. 2008; Kelly and Mapes 2011) and an intercomparison  
11 study of CMIP3 models focusing on the region over Central America and the Intra-America  
12 Seas (Rauscher et al. 2008) concluded that large-scale dynamics such as the NASH do play  
13 an important role in simulating the MSD over the region, similar to observations. More  
14 recent studies based on both observations and model analysis found summer rainfall  
15 variability in the Southeastern United States to be related to the seasonal westward  
16 movement of the NASH, and moreover suggested that recent observed changes in the timing  
17 and position of the NASH that increase the likelihood of both very wet and very dry  
18 summers could be the result of anthropogenic forcing (Li et al. 2011; Li et al. 2012, 2013a).  
19 This result in particular emphasizes the importance of evaluating the GCMs used to generate  
20 future projections of changes in precipitation over regions affected by the NASH, despite  
21 the continuing debate regarding interannual variability and trends in the position of the  
22 NASH and its relationship to global change (Diem et al. 2013; Li et al. 2013b).

23 The bimodal structure of the annual cycle of Caribbean precipitation tends to be  
24 reproduced by Atmospheric Model Intercomparison Project (AMIP) ensembles, in which  
25 SST is prescribed, but not by CMIP ensembles using the same atmospheric models, where  
26 SST is generated internally by the model (Martin and Schumacher 2011). Here, we focus on

1 coupled models to explore the extent to which observed large-scale features such as the  
2 CLLJ, NASH, and SST are reproduced by the models, and how this may impact the ability  
3 of the models to simulate the seasonal cycle of precipitation over the Caribbean.

4 In section 2, we describe the data sources and model outputs used in this analysis.  
5 These include gridded observations, station data, reanalysis, and output fields from the 18  
6 CMIP3 and 26 CMIP5 GCMs. In section 3, we describe the observed climatology of  
7 precipitation over three Caribbean regions (Central America, Greater Antilles, and Lesser  
8 Antilles) and compare observations with model simulations. In section 4, we explore the  
9 large-scale environmental features that drive the annual cycle of precipitation, and assess the  
10 degree to which the GCMs are able to reproduce observed variability. Our results are  
11 summarized and discussed in section 5.

## 12 **2 Data and Model Output**

### 13 **2.1 Study Domain**

14 To characterize climatological precipitation over the Caribbean, we divide the area  
15 into three sub-regions based on availability of station data and geographical location: east,  
16 north-central, and west (Fig. 1). Region 1 encompasses the west side of the Caribbean Sea  
17 and Central America, from the Yucatan Peninsula to the northern border of Panama ( $94^{\circ}\text{W}$   
18  $\sim 83^{\circ}\text{W}$ ,  $8^{\circ}\text{N} \sim 23^{\circ}\text{N}$ ). Region 2 includes most of the larger islands in the central Caribbean,  
19 known as the Greater Antilles: Cuba, Jamaica, Dominican Republic/Haiti, and Puerto Rico  
20 ( $82^{\circ}\text{W} \sim 65^{\circ}\text{W}$ ,  $17^{\circ}\text{N} \sim 25^{\circ}\text{N}$ ). Region 3 includes primarily smaller islands and island  
21 groups located on the east side of the Caribbean, known as the Lesser Antilles: Antigua and  
22 Barbuda, Guadeloupe, Martinique, St. Lucia, etc. ( $64^{\circ}\text{W} \sim 59^{\circ}\text{W}$ ,  $12^{\circ}\text{N} \sim 18^{\circ}\text{N}$ ). The area  
23 of islands in the Lesser Antilles, Region 3, ranges from a few  $\text{km}^2$  to several hundreds of  
24  $\text{km}^2$ . Global model resolution ranges from 1 to 5 degree (Table 1), meaning that the finest

1 horizontal resolution of any model over the region is about 10,000 km<sup>2</sup> (1 degree  $\approx$  100 km).  
2 Clearly, global model simulations are unable to resolve relatively small spatial features such  
3 as the land-sea distribution and the topographically enhanced convective precipitation over  
4 the Lesser Antilles (Region 3). These three domains were chosen to be small enough to  
5 resolve the different characteristics of the annual precipitation cycle across the Caribbean,  
6 yet at the same time large enough to demonstrate a consistent and coherent response to  
7 large-scale dynamics across multiple observational sites.

## 8 **2.2 Observational Data and Reanalysis**

9 Due to data sparsity in this region, we rely on both gridded and station-based  
10 observational data as well as reanalysis. With a relatively small ratio of land-to ocean area in  
11 each study region (see Fig. 1), gridded datasets provide a much-needed broader perspective  
12 on the region as a whole. In contrast, station-based precipitation reflects a more localized  
13 climate that results from island orography but tends to be influenced by small-scale spatial  
14 features which cannot be well-resolved by GCMs (Table 1; see also comparison of gridded  
15 vs. station-based precipitation in section 4).

16 The gridded precipitation data used here comes from the global merged precipitation  
17 data of the Global Precipitation Climatology Project, GPCP version 2 (Adler et al. 2003).  
18 As stated on the project website (<http://www.gewex.org/gpcp.html>), the GPCP combines  
19 data from 6,000 rain gauge stations, satellite geostationary passive microwave instruments,  
20 and sounding observations. This dataset is described as the most complete analysis of  
21 rainfall available over the global oceans, as well as adding additional detail over land areas  
22 with sparse station records, such as the Caribbean. An intercomparison of rainfall products  
23 from observations, reanalysis, satellite and model output for mean rainfall in the Caribbean  
24 (Jury 2009b) showed that although most satellite products and models had a dry bias, the  
25 annual cycle of the reanalysis and satellite products matched observations well. In

1 particular, the GPCP was able to capture key features of annual mean rainfall in the  
2 Caribbean including topographically enhanced convection (central Cuba-Jamaica,  
3 Dominican Republic-Puerto Rico), broad dry zones (northern Lesser Antilles and north of  
4 Venezuela), and a moist tongue (southern Lesser Antilles). GPCP data is gridded to  $2.5^\circ \times$   
5  $2.5^\circ$  in latitude and longitude at monthly resolution from 1979 to 2008 (data available  
6 online at: <ftp://precip.gsgc.nasa.gov/pub/gpcp-v2.1/psg>).

7 For the large-scale dynamical analysis, we use monthly mean geopotential height,  
8 zonal and meridional wind, vertical velocity, and specific humidity from the European  
9 Centre for Medium-Range Weather Forecasts (ECMWF) 40 year reanalysis, ERA-40  
10 (Uppala et al. 2005; data available online at: <http://dss.ucar.edu/datasets/ds119.0>). ERA-40  
11 output is available at a resolution of approximately  $1.4^\circ \times 1.4^\circ$  in latitude and longitude for  
12 the period from September 1957 to August 2002. Reanalysis does not consist of actual  
13 observations, but rather of output from a numerical weather simulation model constrained to  
14 match available observations at regular intervals. In broad terms, reanalysis can be viewed  
15 as informed dynamical interpolation between observations that, for regions such as the  
16 Caribbean, are often relatively sparse.

17 Finally, the SST data used here comes from the National Oceanic & Atmospheric  
18 Administration (NOAA) Optimum Interpolation (OI) SST V2 (Reynolds et al. 2002; data  
19 available at: <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>). SST data  
20 has a spatial resolution of  $1^\circ$  in latitude and longitude and has monthly mean data available  
21 for 27 years, from 1982 to 2008.

22 Two sources of station data for the Caribbean are used to construct monthly time  
23 series for individual locations. The first is 24-hour cumulative precipitation from the Global  
24 Historical Climatology Network (GHCN) database which represents merged station  
25 observations from various sources (data available from the National Climatic Data Center  
26 at: <http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>). The second is sub-daily raw

1 precipitation observations from the U.K. Meteorological Office Integrated Data Archive  
2 System (MIDAS) land surface observational database (data available at:  
3 [http://badc.nerc.ac.uk/view/badc.nerc.ac.uk\\_ATOM\\_dataent\\_ukmo-midas](http://badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_ukmo-midas)), which are  
4 summed to create monthly totals. Both GHCN and MIDAS station data are filtered by a  
5 quality control algorithm. The algorithm tests for repeat (more than 5 consecutive days with  
6 the same reported non-zero value to one tenth of mm) and limits (below 0 mm or above 915  
7 mm, the maximum recorded 24h cumulative precipitation for the Americas). It removes any  
8 instances of these values from the data record and additionally removes any stations with  
9 more than 60% missing data for the 51 years from 1960 to 2010.

10 Fig. 1 also shows the location of the weather stations over the Caribbean region  
11 (between 5°N and 27°N and 94°W and 59°W). Blue dots indicate selected stations, while  
12 light blue dots show stations removed due to data limitations. Grid box centers correspond  
13 to the longitude and latitude of the grid points of the gridded data. To facilitate comparisons  
14 between station and gridded data, station precipitation records are gridded to GPCP  
15 resolution of 2.5° by applying an iterative objective analysis which is a process of  
16 interpolating irregularly spaced data to a fixed grid by successive correction to apply a  
17 relative weight to the data depending on a distance to the grid point (Cressman 1959; Barnes  
18 1964).

### 19 **2.3 Model Information**

20 Eighteen CMIP3 (Meehl et al. 2005) and 26 CMIP5 models (Taylor et al. 2012) that  
21 have archived monthly precipitation data with the Program for Climate Model Diagnosis  
22 and Intercomparison (PCMDI, <http://www-pcmdi.llnl.gov/>) and participated both in phases  
23 3 and 5 of the Coupled Model Intercomparison Project are used in this analysis. A brief  
24 description of the models used in this study, including their names, provenance, and  
25 horizontal and vertical resolution, is provided in Table 1. The three different indices (“B”,

1 “SM” and “S”) next to the model names stand for “Bimodal”, “Single w/MSD”, and  
2 “Single”, respectively. These are the three categories we divided the models into, based on  
3 their performance in simulating the annual cycle of precipitation over Caribbean region 2.  
4 These categories will be discussed in more detail in the next section. For the large-scale  
5 dynamical analysis, we required monthly geopotential height, zonal and meridional wind,  
6 vertical velocity, specific humidity and SST. Not all fields were available for all models;  
7 hence, in Table 1 we also indicate which fields were available for each model, defining the  
8 subset of models on which each analysis (shown in Figures 6 through 9) could be  
9 performed.

10 CMIP3 model simulations for 1960-1999 are driven by the Coupled Model  
11 Intercomparison Project for the 20th Century Climate in Coupled Models (20C3M)  
12 scenarios (Covey et al., 2003). From 2000 to present, model simulations are based on the  
13 Special Report on Emission Scenarios (SRES) mid-high A2 experiment, in which CO<sub>2</sub>  
14 emissions continue to increase, although more slowly than observed over that time  
15 (Nakicenovic and Swart, 2000). CMIP5 model simulations for 1960 to 2005 are driven by  
16 the CMIP5 historical total-forcing simulation (Taylor et al. 2012). From 2006 to present,  
17 model simulations are based on the Representative Concentration Pathway (RCP) mid-low  
18 emission scenario 4.5 (Moss et al., 2010), as this scenario had the largest number of model  
19 simulations and output files available. Future scenarios were used only because “historical”  
20 simulations end in 1999 and 2005 for CMIP3 and CMIP5, respectively, while the data  
21 sources we used ended in 2008 (GPCP) and 2001 (ERA-40). We do not anticipate the  
22 choice of scenario for the last decade will have a significant impact on results, as the  
23 uncertainty in projected trends in precipitation that can be ascribed to different scenarios has  
24 been shown to be negligible over this period at both the global and the regional scale  
25 (Hawkins and Sutton, 2011).

1           Based in part on initial conditions, each simulation establishes its own pattern of  
2 natural variability, including ENSO and Atlantic Multidecadal Oscillation (AMO), etc. such  
3 that no temporal correspondence between simulated and observed conditions at the regional  
4 scale can be expected on a year-by-year basis (Stoner et al. 2009). However, long-term  
5 global average conditions more directly related to global-scale forcing are expected to track  
6 observed trends if the forcing is accurate (e.g., Rahmstorf et al. 2007). Therefore, our  
7 analysis in this study is based on a long-term monthly-mean data, averaged over 30 years  
8 from 1979 to 2008, which is chosen to match the time period of the GPCP (observed) data.

## 9   **3 Comparing Observed and Model-Simulated** 10 **Caribbean Precipitation**

### 11 **3.1 Comparing GPCP to station-based precipitation**

12           The first step in this analysis is to develop a climatology of annual precipitation from  
13 available data. For each of the three regions, we construct composite annual cycles of  
14 precipitation using gridded GPCP monthly data for 30 years from 1979 to 2008 and quality-  
15 controlled station data for the same time period (Fig. 2). The shaded area represents  $\pm 1\text{-}\sigma$   
16 (standard deviation) from the mean. Because of the uneven distribution of the station data  
17 both in time and space, and the fact that most stations are land-based and therefore affected  
18 by small-scale orographic precipitation unresolved at the scale of the gridded GPCP data,  
19 the range of the monthly mean rainfall for the three regions from the stations is much larger  
20 than that from GPCP. Comparing annual cycles of the precipitation based on station data  
21 (Fig. 2, right) with GPCP (Fig. 2, left) reveals both differences and similarities. In terms of  
22 similarities, both datasets demonstrate how the annual cycle of precipitation differs between  
23 the three study regions. Regions 1 (west, Central America) and 2 (central, Greater Antilles)  
24 are characterized by a bimodal distribution, or two wet seasons, with rainfall peaks in June

1 and September (Region 1) and in May and September (Region 2). Region 3 (east, Lesser  
2 Antilles, with the smallest islands and the largest proportion of ocean to land area) shows  
3 the smallest amplitude in the annual precipitation cycle of the three regions, with a wet  
4 season that corresponds to Northern Hemisphere fall.

5         The phase of the annual cycle for Regions 1 (Central America) and 2 (Greater  
6 Antilles) seems to be in a good agreement between GPCP and station data, with the  
7 exception of the second (fall) peak in Region 1, a discrepancy which may be due to the  
8 limited nature of station data. Region 1 stations are concentrated in the northern and  
9 western part of the region (i.e. on the continent of Central America; Fig.1), while GPCP  
10 shows September precipitation increases along the Pacific coast of Central America (not  
11 shown).

12         In terms of other differences, the magnitude of the precipitation in Region 3 (Eastern  
13 Caribbean) derived from station data tends to be much larger than the GPCP data. Region 3  
14 (bounded by 64°W~59°W, 12°N~18°N) is mostly ocean except for several small islands  
15 (see Fig. 1). It is therefore plausible that the sampling issue discussed previously, where  
16 station data more closely reflects rainfall characteristic of small-scale orographic features of  
17 the islands than does a larger gridded dataset, could be strongest for this region. This is  
18 supported by the direction of the difference, where the station data shows higher  
19 precipitation than the gridded dataset. It is well-established that the role of coastal orography  
20 is more often to generate, rather than suppress, local-scale precipitation. In addition, there is  
21 a relatively small amount of station data available in Region 3 as compared to other regions.  
22 This more limited sample might also contribute to disagreement with the GPCP data.

### 23 **3.2 Comparing CMIP3 and CMIP5 simulations to GPCP precipitation**

24         The next step is to compare GPCP-based precipitation with that simulated by CMIP3  
25 and CMIP5 GCMs. GPCP, rather than station-based, climatologies are used due to the

1 sampling biases in station data discussed in the previous section. GPCP includes both  
2 station observations as well as satellite measurements over the land and ocean, and covers a  
3 regular grid similar to that generated by GCMs.

4         Since the Caribbean region is primarily ocean, it is reasonable to expect that SST  
5 exerts a strong influence on precipitation through its effects on atmospheric moisture  
6 content and low-level convergence. A recent study by Martin and Schumacher (2011) found  
7 that Coupled Atmosphere-Ocean (CMIP) models that underestimate SST also tend to  
8 produce less precipitation as compared to Atmosphere-Only (AMIP) models forced by  
9 prescribed (observed) SST. For that reason, before comparing observed and modeled  
10 annual cycle of Caribbean precipitation, we first compare simulated and observed annual  
11 mean precipitation vs. SST for the 30-year historical period 1979-2008 (Fig. 3). Consistent  
12 with Martin and Schumacher (2011), this comparison reveals that all but a handful of  
13 CMIP3 and CMIP5 models systematically underestimate annual mean SST for each of the  
14 three regions. Similarly, the majority of models also underestimate annual mean of  
15 precipitation in Region 1. However, results for Region 2 and 3 tend to be more evenly  
16 distributed. In these regions, models that underestimate SST do not inevitably  
17 underestimate precipitation. As hypothesized by Martin and Schumacher, the sensitivity of  
18 precipitation to SST may be affected by the convective parameterization scheme employed  
19 by different models. This is a topic that remains to be explored in future work.

20         We next compare simulated vs. observed normalized annual cycles of precipitation  
21 for each of the three regions. The normalized annual cycle of precipitation is derived by  
22 subtracting the annual mean from the annual cycle of precipitation and then dividing by the  
23 standard deviation from the annual mean. As such, the normalized annual mean  
24 precipitation is zero. Fig. 4 (CMIP3) and Fig. 5 (CMIP5) show how a clear distinction can  
25 be made between models that do reproduce the bimodal distribution of the wet seasons  
26 identified in the observations for Regions 1 and 2, and those that do not. A similar

1 distinction can be drawn between models that simulate a mid-summer drought, and those  
2 that do not. Based on these distinctions in GCM ability to simulate precipitation for Region  
3 2, where these features are most pronounced, we divide both CMIP3 and CMIP5 GCMs into  
4 three categories as follows.

5         The first category, which we refer to as “Bimodal” (marked as “B” in Table 1),  
6 consists of models that are able to simulate a bimodal distribution similar to that observed in  
7 Region 2, with peaks in early summer and fall punctuated by a drier period in mid-summer.  
8 There are 6 CMIP3 models (CGCM3.1(T47), GISS-ER, MIROC3.2(hires), MRI-  
9 CGCM2.3.2, UKMO-HadCM3, UKMO-HadGEM1) and 9 unique and 13 total CMIP5  
10 models (CanCM4, CESM1(CAM5), CNRM-CM5, CSIRO-Mk3.6.0, HadCM3,  
11 HADGEM2-AO, HADGEM2-CC, HADGEM2-ES, MIROC4h, MIROC5, MPI-ESM-LR,  
12 MPI-ESM-MR, and MRI-CGCM3) in this group. This list of models reveals two interesting  
13 features. First, most of the modeling groups that have an earlier CMIP3 model in this top-  
14 performing category also have a more recent CMIP5 model in the same category. These  
15 consist of the Canadian Centre for Climate Modeling and Analysis, the Japanese Center for  
16 Climate System Research, the Japanese Meteorological Agency, and the U.K. Met Office  
17 Hadley Centre. Second, there appears to be some improvement between CMIP3 and CMIP5  
18 (see Table 1 and Fig. 4 and 5). One model from the Australian Commonwealth Scientific  
19 and Industrial Research Organization (CSIRO-Mk3.0, CSIRO-Mk3.6.0) moves from  
20 category 3 (single) in CMIP3 to category 1 (bimodal) in CMIP5, while one from the  
21 German Max-Planck-Institut für Meteorologie moves from category 2 (single with MSD) in  
22 CMIP3 (ECHAM5/MPI-OM) to category 1 in CMIP5 (MPI-ESM-LR, MPI-ESM-MR).  
23 One model from the U.S. National Center for Atmospheric Research (CESM1(CAM5)) also  
24 debuts in category 1.

25         The second category, which we refer to as “single with MSD” (marked as “SM” in  
26 Table 1), consists of models that simulate both a mid-summer drought and a wet season in

1 fall, but fail to simulate the early summer precipitation maxima. There are 6 CMIP3 models  
2 (GFDL-CM2.0, GFDL-CM2.1, GISS-EH, INM-CM3.0, MIROC3.2(medres) and MPI-  
3 ECHAM5) and 3 unique and 7 total CMIP5 models (CCSM4, CESM1(BGC),  
4 CESM1(WACCM), GFDL-ESM2G, GFDL-ESM2M, MIROC-ESM, MIROC-ESM-  
5 CHEM) in this group. Some models remain in this category from CMIP3 to CMIP5,  
6 specifically the CMIP5 Earth System Model versions of CMIP3 models from the U.S.  
7 Geophysical Fluid Dynamics Laboratory (GFDL-ESM2G and GFDL-ESM2M, plus GFDL-  
8 CM2.0 and GFDL-CM2.1). Models from the Japanese Center for Climate System Research  
9 remain in the same category, depending on the horizontal resolution of the model. This  
10 appears to be independent of model version for both CMIP3 and CMIP5. Specifically,  
11 models with a coarse resolution of about  $2.8^\circ$  in latitude and longitude (MIROC-ESM and  
12 MIROC-ESM-CHEM, plus MIROC3.2(medres)) are classified as category 2, while models  
13 with a higher resolution less than about  $1.4^\circ$  (MIROC4h, MIROC5, plus MIROC3.2(hires))  
14 fall into category 1. Some models have moved into category 2 from category 3, which  
15 represents an improvement from CMIP3 to CMIP5. These include the majority of the  
16 CMIP5 versions of the U.S. National Center for Atmospheric Research models (CCSM4,  
17 CESM1(BGC), CESM1(WACCM), compared to CCSM3 and PCM which are in the  
18 “single” category).

19 Finally, the third category, which we refer to as “single” (marked as “S” in Table 1),  
20 consists of models that simulate only a single wet season from June to October, with no  
21 peak in early summer and no mid-summer drought. There are 5 CMIP3 models  
22 (CGCM3.1(T63), CSIRO-Mk3.0, CCSM3, GISS-AOM and PCM) and 3 unique and 6 total  
23 CMIP5 models (GFDL-CM3, GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC,  
24 INM-CM4) in this group. Interestingly, most CMIP5 models in this group have regressed  
25 from category 2 (single with MSD) into category 3 (single only) between CMIP3 and  
26 CMIP5. These include the new CMIP5 GFDL-CM3 (compared to GFDL-CM2.0 and 2.1),

1 many of GISS models (GISS-E2-H-CC, GISS-E2-H, GISS-E2-R-CC, and GISS-E2-R,  
2 compared to GISS-EH), and the new CMIP5 INM-CM4 (compared to INM-CM3.0). Three  
3 modeling groups (Canadian, Australian, and U.S. NCAR) have moved out of category 3  
4 between CMIP3 and CMIP5.

5 One CMIP3 model, CNRM-CM3, could not be assigned to any of these three  
6 categories, as its annual precipitation profile did not correspond to any of the general  
7 profiles (not shown). From March to October, encompassing the summer wet season and the  
8 mid-summer drought, the model would fall into the “Bimodal” category. However, the  
9 model also produced a spurious winter wet season, extending the fall precipitation peak  
10 throughout the entire winter, from September through January. In Table 1, CNRM-CM3 is  
11 therefore categorized as “N/A”. In CMIP5, the more recent version of the model, CNRM-  
12 CM5 eliminated the spurious winter precipitation peak and is now one of the top performers  
13 relative to the multi-model ensemble of month-to-month precipitation.

14 In summary, CMIP5 versions of models that were already able to simulate a bimodal  
15 distribution for Regions 2 in CMIP3 tend to retain that ability with the new generations of  
16 models. The majority of modeling groups show improvement in the ability to simulate the  
17 annual cycle of precipitation over the Caribbean, moving from “single” to “single with  
18 MSD” or “bimodal” between CMIP3 and CMIP5, while a few models appear to have  
19 regressed in their ability to simulate the annual cycle of precipitation from CMIP3 to  
20 CMIP5.

21 The discussion above is based on model performance over Region 2. In general,  
22 model performance in other regions is consistent with performance in Region 2. However,  
23 there are a handful of exceptions. Specifically, three CMIP3 models (GISS-ER, which has a  
24 bimodal distribution in Region 2 but not Region 1; and CSIRO-Mk3.0 and MPI-ECHAM5,  
25 both of which have a bimodal distribution in Region 1 but not Region 2) and one CMIP5  
26 model (CESM1(CAM5), which does not have a bimodal distribution in Region 1) would fall

1 in a different group if they were divided by their performance over Region 1 instead. The  
2 majority of models are able to simulate a wet season in fall in Region 3 regardless of their  
3 performance in Regions 1 and 2. The only exception are models in category 3, “single”,  
4 which tend to simulate a Region 3 wet season in summer instead of fall.

5 In the next section, we use these categories of models to explore the association  
6 between large-scale environmental features, such as SST, NASH, and CLLJ, and the annual  
7 cycle of precipitation over the Caribbean. For those models with archived upper-air fields,  
8 we also examine the extent to which aspects of these large-scale features are simulated by  
9 the GCMs.

## 11 **4 Influence of the Large-Scale Environment on** 12 **Precipitation**

13 In evaluating the role of the large-scale environment on precipitation in the  
14 Caribbean, the question we ask is not whether all the models are able to reproduce the  
15 observed large-scale environment, but rather whether there is a noticeable difference among  
16 the performance of the models in the “bimodal” category (which correctly simulate the  
17 annual cycle of precipitation with two maxima and a mid-summer drought), models in the  
18 “single with MSD” category (which simulate a mid-summer drought and fall precipitation  
19 maxima) and models in the “single” category (which only simulate a single precipitation  
20 maxima, generally in summer). In this section we continue to divide the models into these  
21 three categories, and examine their ability to reproduce monthly changes in SST, the NASH,  
22 low-level wind fields including the CLLJ, and moisture flux convergence throughout the  
23 region. Not all models have archived the fields required for these analyses; for each figure,  
24 available models are listed in Table 1.

## 1 **4.1 Sea Surface Temperature**

2 As previous studies have already demonstrated how sea surface temperature plays an  
3 important role in driving the annual cycle of climate in the Caribbean (Enfield and Alfaro  
4 1999, Giannini et al. 2000, Munoz et al. 2008), we start with examining month-to-month  
5 variation in the SST across the region spanning the wet season from May to September (Fig.  
6 6, shaded area). As shown in the observations (left column), a pool of warm SST first  
7 appears along the Pacific coast of Central America in April. It then intensifies and begins to  
8 shift northeastward into the Caribbean in May. SST near Cuba begins to increase in May  
9 (Region 2) and spreads to cover the entire region by June. During summer, the warm SST  
10 anomaly extends throughout the entire Gulf of Mexico and northern Caribbean, covering a  
11 latitude band between about 18°N and 30°N in August, before rapidly decreasing by  
12 October.

13 CMIP3 simulations (Fig. 6, right-hand columns) show how most models tend to  
14 under-estimate SST relative to observations. This is consistent with the annual SST analysis  
15 in Fig. 3 discussed previously. Most models also show a slight delay in the advance of the  
16 SST warm pool into the Caribbean in spring. As hypothesized, the multi-model average of  
17 the first or Bimodal category of CMIP3 models (CGCM3.1(T47), GISS-ER,  
18 MIROC3.2(hires), MRI-CGCM3.2, UKMO-HadCM3, and UKMO-HadGEM1) is most  
19 similar to observed (Fig. 6, second column). Compared to the other categories and to  
20 observations, the multi-model average of category 2 CMIP3 model (ECHAM5/MPI-OM,  
21 GISS-EH, GFDL-CM2.0, GFDL-CM2.1, INM-CM3.0, and MIROC3.2(medres)) displays a  
22 stronger meridional gradient of SST across the Caribbean due to a colder pool of SST  
23 located near the northern coast of South America (Fig. 6, third column). The three-model  
24 average of category 3 CMIP3 models (CSIRO-Mk3.0, GISS-AOM, and PCM) shows a  
25 delayed and weaker SST than both the observed and the other categories of model-simulated  
26 SST.

1           The same three categories of multi-model averages, this time based on CMIP5  
2 simulations, are plotted in Fig. 7 to determine whether the relationship between model  
3 category and SST is consistent with that seen in CMIP3 simulations (Fig. 6). The most  
4 noticeable difference between Fig. 6 and 7 is that the negative or cool bias in model-  
5 simulated SST across the Caribbean and the Gulf of Mexico seems to be reduced in CMIP5  
6 simulations compared to CMIP3. Consistent with the CMIP3 simulations, the multi-model  
7 average of SST for the first Bimodal category (CanCM4, CESM1-CAM5, CNRM-CM5,  
8 CSIRO-Mk3-6-0, HadCM3, HadGEM2-AO, HadGEM-CC, HadGEM-ES, MIROC4h,  
9 MIROC5, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3) is most similar to observations on  
10 a month-to-month basis. The average of the second “Single with MSD” category (CCSM4,  
11 CESM1-BGC, CESM1-WACCM, GFDL-ESM2G, GFDL-ESM2M, MIROC-ESM-CHEM,  
12 MIROC-ESM) displays a similar month-to-month variation to category 1 with the exception  
13 of a cold SST bias along the northern coast of South America in the early summer (May and  
14 June). This is also consistent with the CMIP3 simulations, strengthening our confidence in  
15 the similarity of model biases between CMIP3 and CMIP5. In the CMIP5 simulations, it is  
16 very clear that the warm SST pool in the Gulf of Mexico persists into September, when SST  
17 in the observations and in the bimodal simulations has already begun to decrease. This late-  
18 season bias is not as evident in the CMIP3 simulations, but is nonetheless present upon  
19 careful inspection. The third or “Single” category of CMIP5 models (GFDL-CM3, GISS-  
20 E2-H, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC, INM-CM4) also has an early-season  
21 cool bias off the northern coast of South America, but otherwise shows a somewhat different  
22 spatial distribution of SST as compared to CMIP3. Instead of a northward propagation of the  
23 warm SST pool over both the Pacific and Atlantic Oceans during the wet season, the  
24 warmest SST remains to the south of 14°N in the eastern Pacific while the second warmest  
25 region appears to be near the central Caribbean rather than the the Gulf of Mexico. This  
26 may be due to the fact that the models in the CMIP5 category 3 are not updated versions of

1 the models in the CMIP3 category 3, but rather entirely different models (with the sole  
2 exception of models from the same group, GISS).

3 The timing of the month-to-month variation in SST suggests that it could play a role  
4 in moderating the onset and end of the wet season from May to October (June to October) in  
5 Region 2 (Region 1). The timing of the onset of precipitation maxima in these regions in  
6 spring is similar to that of the intrusion of the warm SST pool into those regions, and the  
7 decrease in precipitation in fall also corresponds to the retreat of the warm SST pool from  
8 those areas. Most models (particularly in the single with MSD and single categories) tend to  
9 simulate a wet season that is slightly delayed compared to observations (Fig. 4 and 5),  
10 consistent with a delay in simulating the arrival of the SST warm pool. On the other hand,  
11 precipitation in Regions 1 and 2 decreases during July and August, just when the SST in  
12 both regions is strongest. This suggests that atmospheric features, such as the NASH and  
13 CLLJ, may be more relevant to determining the existence and timing of the mid-summer dry  
14 period than SST (as previously suggested by Curtis and Gamble 2007; Gamble et al. 2008  
15 and Rauscher et al. 2008). In other words, the midsummer dry period seen over the central  
16 and eastern Caribbean (illustrated in Fig. 4 and 5) is thought to be related to an enhanced  
17 CLLJ accompanied by westward expansion of the NASH over the Caribbean, which will be  
18 discussed in the next section.

## 19 **4.2 The Caribbean Low-Level Jet and Geopotential Height**

20 In Fig. 6 and 7 (left column) we overlay ERA40-based climatological zonal wind  
21 (blue contours) and eddy geopotential height at 925hPa (black contours) over SST in order  
22 to track the extension of the NASH and associated CLLJ from the tropical Atlantic (in  
23 spring) into the Caribbean (in summer), and its retreat back to the east (in fall). We derived  
24 eddy geopotential height by subtracting the zonal mean from the geopotential height at each  
25 pressure level. The eddy refers to deviation from zonal mean and the units of this derived

1 variable are in meters relative to the zonal mean. Since the zonal mean field is zonally  
2 symmetric, the eddy geopotential height can be used to track the westward extension of the  
3 NASH. Specifically, geopotential height maxima highlight the extension of the NASH into  
4 the Caribbean in summer, while the easterly zonal wind maxima is the CLLJ. The CLLJ is  
5 located to south of the eddy geopotential height, parallel to the isobars thus maintaining  
6 geostrophic balance with the NASH (Wang 2007; Cook and Vizy 2010).

7         Towards the end of Northern Hemisphere spring, the CLLJ begins to increase in  
8 strength over the Caribbean Sea and remains strong from June through August before  
9 weakening again in fall. These increased easterly winds, centered around 15°N, seem to be  
10 associated with the increase of the SST over the Gulf of Mexico and the Caribbean Sea  
11 between 18°N and 30°N. This makes sense, since an increase in the meridional gradient of  
12 SST would tend to strengthen easterly winds via geostrophic balance. At the same time, the  
13 NASH begins to extend westward, stalling over the Caribbean and the Gulf of Mexico  
14 during July and August before retreating eastward in September to the point where the  
15 geopotential height anomaly over the Caribbean is nearly non-existent in October.

16         As shown previously in Section 3, GCM ability to simulate the mid-summer drought  
17 varies noticeably between the three categories of models identified here. The first two  
18 categories (bimodal and single with MSD) are able to simulate a drought, while the last one  
19 (single) is not. For that reason, we next examine how the ability of the models to simulate  
20 the mid-summer dry period and the timing of the wet season peak in the western and central  
21 Caribbean may be related to the timing of the westward extension and eastward retreat of  
22 the NASH and/or the intensification of the CLLJ. Specifically, we hypothesize that a too-  
23 early extension of the NASH into the Caribbean would suppress the early summer  
24 precipitation maxima, whereas a model that simulated the proper timing of NASH extension  
25 would allow for an SST-driven increase in precipitation in early summer. On the other hand,  
26 model inability to simulate the extension of the NASH over the Caribbean would mean that

1 the model was not able to simulate the MSD, as there would be no atmospheric mechanism  
2 to suppress SST-driven precipitation.

3 In accord with our hypotheses, both Fig. 6 (CMIP3) and Fig. 7 (CMIP5) show that  
4 the multi-model average of the bimodal category is able to simulate the timing of the  
5 westward extension of the NASH over the eastern Caribbean and the Gulf of Mexico and  
6 the accompanied intensification of the CLLJ during July and August. Both CMIP3 and  
7 CMIP5 model simulations in the Single with MSD category, which simulates only a single  
8 wet season in fall, tends to simulate an early westward extension of the NASH that expands  
9 over the central Caribbean in May and over the Gulf of Mexico in June. In addition, the  
10 early summer CLLJ is also much stronger than observed, followed by the weakening of the  
11 CLLJ in mid-summer (July and August). On the other hand, the multi-model average of the  
12 Single category shows a weak CLLJ throughout the wet season and a relatively weak  
13 westward extension of the NASH over the Caribbean, more prominent in the CMIP5  
14 simulations but present in CMIP3 as well. In summary, the combined effects of the  
15 longitudinal migration of the NASH, the accompanying intensification of the CLLJ, and the  
16 annual cycle of SST in the Caribbean appear to drive the annual cycle of the precipitation  
17 over the Caribbean. Decreases in precipitation over the western and central Caribbean  
18 during mid-summer seem to be related to the westward extension of the NASH  
19 accompanying an enhanced CLLJ, which is in turn associated with an increase in the  
20 meridional gradient of SST. As the NASH retreats eastward in the fall, precipitation peaks  
21 again in the western and central Caribbean as nearby SST remains warm, favorable for  
22 convective activity. In the eastern Caribbean, however, the NASH moves in before SST has  
23 warmed in early summer. SST then peaks during September and October (Fig. 2) just as the  
24 NASH retreats to the east. This explains the single precipitation peak in Region 3, the  
25 Lesser Antilles. This result is consistent with Curtis and Gamble (2007) who showed that  
26 the occurrence of maximum of 1000 hPa geopotential height and the MSD progressed from

1 east to west, while the MSD is not observed in the eastern Caribbean. Consistency in multi-  
2 model composite SST, geopotential height, and zonal wind fields suggests that model ability  
3 to simulate the timing and magnitude of these features appears to be related to model ability  
4 to simulate the early-season precipitation maxima (suppressed if the NASH moves westward  
5 too early) and the mid-summer drought (does not occur if the NASH is too weak). Since the  
6 CMIP3 and CMIP5 overall agree on the relationship between model ability to simulate the  
7 annual cycle in Caribbean precipitation and model ability to simulate the large-scale features  
8 such as the NASH, CLLJ and SST, and since CMIP3 model output for upper-air fields is  
9 significantly more sparse than CMIP5 output, in the next section we rely on CMIP5  
10 simulations only to explore the role of vertical wind and moisture convergence/divergence  
11 in determining summer precipitation over the Caribbean.

### 12 **4.3 Exploring Top-Down vs. Bottom-Up Influences on Caribbean** 13 **Precipitation**

14 The influence of oceanic (SST) vs. atmospheric (NASH, CLLJ) influences on  
15 Caribbean precipitation can be difficult to untangle, as the westward extension of the NASH  
16 occurs at a similar time to the eastward extension of the warm SST pool into the Caribbean  
17 Sea and the Gulf of Mexico. In fact, it may be a chicken and egg problem, as SST cooling  
18 south of the NASH, where the CLLJ is most intense, might contribute to maintenance of the  
19 geostrophic balance between the NASH and the CLLJ. For that reason, in Fig. 8 we further  
20 investigate these relationships by plotting latitude-pressure cross-sections of eddy  
21 geopotential height, easterly winds, and local meridional circulation averaged from 80° W to  
22 60°W. During boreal winter, the CLLJ appears to be strongest below 700 hPa where the  
23 southward branch of a local Hadley circulation exists. In contrast, a positive eddy  
24 geopotential height is strongest in the upper troposphere above the Caribbean (not shown).  
25 During summer, observations (Fig. 8, left column) show that the positive eddy geopotential

1 height above the Caribbean is strongest near the surface. Considering the NASH is a  
2 dynamical feature dominating in the lower troposphere (Davis et al. 1997; Li et al. 2011; Li  
3 et al 2012), the enhancement of the positive eddy geopotential height in the lower  
4 troposphere during summer seems to be related to the extension of the NASH over the  
5 Caribbean, presumably indicating that the positive eddy geopotential height is more likely  
6 being affected by the surface rather than the upper atmosphere. In other words, the warming  
7 of the SST in the Caribbean and the Gulf of the Mexico and the cooling of the SST south of  
8 the NASH associated with the enhanced CLLJ drives a positive meridional gradient in near-  
9 surface temperature. This gradient could conceivably play a role in the westward extension  
10 of the NASH by maintaining of the geostrophic balance between the NASH and the CLLJ.  
11 This hypothesis is supported by the observation that the CLLJ appears to be deeper during  
12 boreal summer than winter, expanding up to 500 hPa during July and August. Consistent  
13 with the geostrophic relationship between the CLLJ and the NASH, the height of the CLLJ  
14 matches the height of the positive eddy geopotential height.

15         The observed vertical structures of the CLLJ and eddy geopotential height are both  
16 well simulated by models in the first Bimodal category (Fig. 8, column 2). Consistent with  
17 Fig. 7, the CLLJ and the eddy geopotential height, as well as local meridional circulation,  
18 appear to be stronger in the second Single with MSD category (Fig. 8, column 3).  
19 Specifically, a strong sinking motion above the Caribbean ( $10^{\circ}\text{N} \sim 30^{\circ}\text{N}$ ) seems to be  
20 relevant to the stalling of the strong positive eddy geopotential height (i.e. high pressure  
21 system) during summer and moreover helps to inhibit convection activity, which  
22 consequently prevents the models from simulating a wet season in early summer. As  
23 expected from Fig. 7, the strength of the CLLJ and eddy geopotential height are both  
24 underestimated in the single category (Fig. 8, last column), which indicates that the models  
25 of this category are generally unable to simulate the intensification of the NASH over the  
26 Caribbean during summer.

1           According to Martin and Schumacher (2011), the bimodal structure of the annual  
2 cycle of the Caribbean precipitation tends to be realized by ensembles of AMIP simulations,  
3 but not by ensembles of CMIP simulations. Our intercomparison of the CMIP3 and CMIP5  
4 GCMs with observations identifies a number of models (those in the bimodal category) that  
5 are able to correctly time the westward expansion of the NASH and hence are able to  
6 correctly simulate the observed bimodal distribution of Caribbean precipitation. In models  
7 that anticipate the westward extension of the NASH, however, this first peak is suppressed,  
8 leading to the category “Single with MSD”. On the other hand, models that fail to simulate  
9 the westward extension of the NASH and the enhanced CLLJ, a failure associated with a  
10 weak meridional gradient of SST across the Caribbean and the Gulf of Mexico, also tend to  
11 simulate a wet season during boreal summer, often at the same time as the observed MSD  
12 occurs. This leads to the category “Single”. For CMIP3 models, roughly each one-third of  
13 all models fall into each category. Hence, it is possible that averaging across all CMIP3  
14 models may have obscured the ability of some models to reproduce the observed structure  
15 of precipitation.

#### 16 **4.4 Vertically integrated moisture flux convergence**

17           Previous studies have pointed out that the enhanced CLLJ could be responsible for a  
18 decrease in precipitation in the Caribbean through inhibition of convective development and  
19 tropical cyclogenesis due to divergence over the eastern Caribbean (Wang and Enfield  
20 2003; Wang 2007) or SST cooling due to strong wind stress (Giannini et al. 2000).  
21 However, Regions 1 and 2, which have a mid-summer drought in July and August, are  
22 located to the north of the CLLJ (see Fig. 1 for region locations) where the nearby SST is  
23 still warm during time of drought (Fig. 6, left column). In the area where the CLLJ is  
24 strong, SST does tend to be cooler, suggesting that surface wind stress due to the CLLJ may  
25 play a role in cooling SST in the southern Caribbean Sea, where the CLLJ is strongest.

1 Positive feedback between the CLLJ and the meridional gradient of SST could in part  
2 contribute to the intensification of the NASH over the Caribbean region to the north of the  
3 CLLJ. Large-scale subsidence that occurs within a high pressure system is an unfavorable  
4 condition for convective development or tropical cyclogenesis, processes responsible for  
5 much of the precipitation in Caribbean regions. In that way, SST cooling associated with  
6 the enhanced CLLJ might indirectly influence Caribbean precipitation. Model composites  
7 for both the first category, Bimodal, and the second category, Single with MSD, are able to  
8 simulate cooler SST in the region of strongest CLLJ (Fig. 6 and 7, centre columns).

9 To examine the potential dynamical linkage of the westward extension of the NASH  
10 with Caribbean precipitation, we next examine vertically integrated moisture flux  
11 divergence and horizontal wind at 925 hPa (Fig. 9). Originally, we calculated both  
12 climatological monthly mean and transient moisture flux using 6-hourly output from ERA-  
13 40 reanalysis from 1960 to 2001. Since the total vertically integrated moisture flux  
14 convergence for each month is dominated by the mean term, however (not shown), in this  
15 figure we show only the vertically integrated moisture flux convergence calculated from  
16 monthly mean data.

17 As expected, moisture flux convergence dominates in Region 1 over Central  
18 America and Region 2 over the Greater Antilles and central Caribbean (Region 2) during the  
19 two wet seasons, May/June and September (Fig. 9, left column). In June, moisture flux  
20 divergence is located over Region 3, the eastern Caribbean. Over the summer, this  
21 divergence moves westward. By July and August, it covers most of the Caribbean, the Gulf  
22 of Mexico, and a part of Central America. This corresponds well to the timing of the  
23 westward extension of the NASH over the same period (Fig. 6 and 7) and the mid-summer  
24 drought in Regions 1 and 2. Wind flow during summer is dominated by eastward flow in  
25 the southern Caribbean and an anticyclonic flow pattern in the northern part of the region

1 that is strongest in the west, connecting the Caribbean to the northerly Great Plains Low-  
2 Level Jet.

3         Month-to-month variations in the vertically integrated moisture flux appears to be  
4 well simulated by models in the first Bimodal category, consistent with their ability to  
5 simulate the timing and magnitude of the westward extension and retreat of the NASH. In  
6 the second Single with MSD model category, however, vertically integrated moisture flux  
7 divergence dominates over the Caribbean and the Gulf of Mexico during the entire boreal  
8 summer from May to August. This corresponds to the early expansion of the NASH over the  
9 Caribbean and suppression of the first precipitation maxima in spring. It indicates that the  
10 failure of these models to simulate the first wet season in May/June can be attributed to the  
11 moisture flux divergence associated with the NASH extending too far to the west during  
12 early summer. Finally, vertically integrated moisture flux convergence/divergence  
13 simulated by the models with only a single precipitation maxima shows moisture flux  
14 convergence to be dominant over Central America (Region 1) and most of the Greater  
15 Antilles (Region 2) even in July and August. The pattern over much of the summer  
16 resembles a weaker version of the observed June pattern, which characterizes the first wet  
17 season in the Caribbean. The persistence of this June-like pattern over the entire summer  
18 can be attributed to model inability to simulate a strong NASH extension over the region in  
19 summer, which also results in a weak CLLJ and weak anticyclonic circulation. It results in  
20 the single wet season, beginning in June, that characterizes models in this group.

## 21 **5 Discussion and Conclusions**

22         In this study, we assessed the ability of CMIP3 and CMIP5 GCMs to simulate  
23 observed precipitation in the Caribbean region and linked their performance to model ability  
24 to simulate larger-scale regional dynamics related to SST, NASH, and the CLLJ. We  
25 divided the Caribbean up into three regions based on the groupings of the land mass and the

1 unique characteristics of their annual precipitation: Caribbean Central America, Central  
2 Caribbean (Greater Antilles), and Eastern Caribbean (Lesser Antilles). We then  
3 characterized the annual climatology of precipitation in each of these regions by comparing  
4 the results from the GPCP gridded dataset with those obtained from spatial interpolation of  
5 GHCN and MIDAS station data.

6 Station-based and gridded datasets were consistent in revealing inter-regional  
7 differences in the timing and magnitude of precipitation variability over the year.  
8 Differences between station-based vs. gridded cycles are consistent with the disparate  
9 land/ocean sampling between the datasets, although it is possible that the greater amount of  
10 missing data in the station records as compared to the gridded datasets may also play a role.

11 Region 3 (Eastern Caribbean) is characterized by a single fall wet season that occurs  
12 after the NASH retreats eastward. In contrast, the wet season in Regions 1 (Central  
13 America) and 2 (Central Caribbean) spans the period from May to October, and is  
14 characterized by a bimodal distribution punctuated by a mid-summer drought in July and  
15 August. The overall duration of the wet season is closely linked to the expansion of a warm  
16 SST anomaly into those regions. However, the incidence of the mid-summer dry period is  
17 coincident with the warmest SSTs, suggesting that the MSD may be driven by atmospheric  
18 rather than ocean conditions. The most likely candidate is the westward extension of high  
19 pressure conditions associated with the North Atlantic Subtropical High into the Caribbean,  
20 as revealed by an analysis of 925 hPa geopotential height fields. This high pressure field  
21 may be further enhanced by SST increases throughout the Gulf of Mexico and the Atlantic  
22 Ocean in summer months, which cause the meridional gradient of SST across the Central  
23 Caribbean to peak in July and August. This gradient drives a geostrophic enhancement in  
24 the Caribbean Low-Level Jet, which accompanies the intensification of the NASH over the  
25 Caribbean region to the north of the CLLJ.

1           A vertical cross-section of ERA-40 reanalysis averaged over the Caribbean region  
2 revealed that the CLLJ gradually extends up to about 500 hPa over the course of the summer  
3 rainy season, with a maxima near the surface during July and August concurrent with and  
4 enhancement of eddy geopotential height from the surface. This result supports our  
5 hypothesis that the northward migration of the warm SST anomaly toward the central  
6 Caribbean and the Gulf of Mexico during summer helps the NASH to expand to the west,  
7 and that the intensification of the CLLJ, accompanied by the westward migration of the  
8 NASH (due to the easterly of the anticyclonic circulation), presumably induces the SST  
9 cooling near the northern coast of the South America as a result of wind-driven upwelling.  
10 As a result, the meridional gradient of the SST across the region increases, which in turn  
11 results in both the enhancement of the CLLJ and intensification of the NASH in the  
12 geostrophic relationship. When the warm SST anomaly decreases and retreats to the south  
13 in September, both the CLLJ and the NASH are noticeably weakened. Moreover, the  
14 intensified NASH in the lower troposphere (below 500 hPa), where most of the tropospheric  
15 moisture content exists, helps to inhibit convection development by causing large-scale  
16 moisture flux divergence, particularly during July and August. In summary, Caribbean  
17 precipitation appears to be influenced by the competing effects of the eastward extension of  
18 the warm SST pool in summer which increases precipitation, and the westward migration of  
19 the North Atlantic Subtropical High and associated strengthening of the Caribbean Low  
20 Level Jet and the meridional SST gradient, which suppresses precipitation.

21           The observed cycle in annual precipitation based on GPCP was then compared with  
22 those simulated by 18 CMIP3 GCMs and 26 CMIP5 GCMs. GCM ability to simulate the  
23 annual cycle of precipitation over the Caribbean appears to be linked to model ability to  
24 simulate the large-scale features driving precipitation, which are relatively complex. It is  
25 therefore no surprise that GCM performance varies noticeably, in ways that appear related  
26 to model ability to simulate the timing of changes in multiple dynamical features over the

1 Caribbean. In this analysis, we grouped models into three categories based on their ability to  
2 simulate observed precipitation, then examined the large-scale environment in these models  
3 to see whether consistent patterns emerged.

4 We first found that models able to realistically capture summer changes in the  
5 meridional gradient of SST and the westward extension of the NASH as manifested in 925  
6 hPa geopotential height fields are also able to simulate the bimodal distribution of the  
7 annual cycle of the Caribbean precipitation: a wet season stretching from late spring to early  
8 fall, punctuated by a mid-summer dry period. Moreover, the month-to-month variations of  
9 the vertical structure of the CLLJ and eddy geopotential height (i.e. NASH) are very similar  
10 to observed, indicating that our hypothesis regarding large-scale drivers of precipitation  
11 explained above works well in this model group. Similarly, the ability of these models to  
12 simulate the observed mid-summer dry period can be attributed to mid-summer large-scale  
13 moisture flux divergence due to the westward extension and intensification of the NASH.  
14 These models were designated as “Bimodal” and included approximately one-third of  
15 CMIP3 models but approximately half of all CMIP5 models, suggesting that some modeling  
16 groups' ability to simulate the large-scale drivers of precipitation over the Caribbean has  
17 improved from CMIP3 to CMIP5.

18 We next identified a group of models that simulated an earlier extension of the  
19 NASH into the Caribbean region combined with colder SSTs, which tends to suppress the  
20 early wet season in May and June, but still reproduces the mid-summer dry period.  
21 Consistent with the earlier extension of the NASH, a stronger CLLJ is simulated in the early  
22 summer although the NASH still stalls over the Caribbean and the Gulf of Mexico in mid-  
23 summer. Based on this analysis, it is possible that early onset of the NASH over the  
24 Caribbean might be related to a dominant subsidence throughout the troposphere above the  
25 Caribbean, which would suppress precipitation. This can be anticipated from the fact that  
26 large-scale moisture flux divergence dominated across the region during the wet season

1 from May to August. These models were designated as “Single with MSD” and included  
2 approximately one-third of CMIP3 models and about a quarter of CMIP5 models.

3 Lastly, we identified a group of models that did not simulate a mid-season dry period  
4 in Regions 1 and 2, but rather a single, long wet season which began around June and lasted  
5 until fall. These models were characterized by remarkably weak CLLJ and NASH over the  
6 Caribbean, most prominent in CMIP5 simulations. Consistently, these models tend to  
7 reproduce observed June-like patterns of moisture convergence in the Central America  
8 (Region 1) and the central Caribbean (Region 2) over the entire summer from June through  
9 September. This presumably contributes to the absence of the mid-summer dry period.  
10 These models were designated as “Single” wet season and included approximately one-third  
11 of CMIP3 models but only about a quarter of CMIP5 models.

12 Overall, our result suggests that some models' ability to simulate the large-scale  
13 drivers of precipitation over the Caribbean has improved from CMIP3 to CMIP5. In other  
14 words, more models are now able to reproduce the bimodal structure of the annual cycle of  
15 Caribbean precipitation in the CMIP5 simulations as compared to CMIP3 (see Table 1).  
16 However, we also found that many of the “Bimodal” and “Single with MSD” models from  
17 both the CMIP3 and CMIP5 simulations continue to underestimate the magnitude of the  
18 early wet season, while over-estimating the second wet season in fall. This pattern is more  
19 obvious in the Region 2, central Caribbean (Fig. 4 and 5). This may be due to an  
20 approximate one-month delay in the onset of warm SST across the Caribbean that is  
21 generally seen in nearly all models compared to observations (Fig. 6 and 7).

22 It is possible that the SST bias simulated by the GCMs could be induced by  
23 atmospheric dynamics, perhaps due to biases in wind stress related to wind-driven upwelling  
24 or evaporative cooling. It is also possible that the biases could be induced by ocean  
25 dynamics. In particular, the ocean mixed layer depth is closely related to SST variability  
26 and SST response to atmospheric forcing (Houghton 1991; Yu et al. 2006; Lienert et al.

1 2011). If the simulated depth of the ocean mixed layer were greater than observed, greater  
2 ocean heat content would increase the inertia or delay in the response of surface SST to  
3 seasonal heating. These and other hypotheses regarding the origin of the observed biases in  
4 model simulation of the large-scale atmospheric and oceanic environment remain to be  
5 explored in future research, as do the implications of GCM ability to simulate precipitation-  
6 related dynamics on the magnitude and direction of future change in the region.

## 8 **Acknowledgements**

9 The authors would like to thank two anonymous reviewers for constructive suggestions that  
10 improved this manuscript. We also thank Mr. Caleb Crow for quality control of station data.  
11 We acknowledge the World Climate Research Programme's Working Group on Coupled  
12 Modelling, which is responsible for CMIP, and we thank the climate modeling groups  
13 (listed in Table 1 of this paper) for producing and making available their model output. For  
14 CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and  
15 Intercomparison provides coordinating support and led development of software  
16 infrastructure in partnership with the Global Organization for Earth System Science Portals.  
17 This research was supported by USGS Award Number G10AC00582.

## 1 **References**

- 2
- 3 Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U,  
4 Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, 2003: The Version 2 Global  
5 Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-  
6 present). *J Hydrometeorol* 4:1147-1167.
- 7 Amador J (1998) A climatic feature of the tropical Americas: The trade wind easterly jet.  
8 *Top Meteor Oceanogr* 5:91-102.
- 9 Barnes SL (2007) A Technique for Maximizing Details in Numerical Weather Map  
10 Analysis. *J Appl Meteorol* 3:396-409.
- 11 Cressman, GP (1959) An Operational Objective Analysis System. *M Weather Rev* 87:367-  
12 374.
- 13 Cook KH and Vizy EK (2010) Hydrodynamics of the Caribbean Low-Level Jet and Its  
14 Relationship to Precipitation. *J Climate* 23:1477-1494.
- 15 Covey C, AchutaRao KM, Cubasch U, Jones P, Lambert S, Mann M, Phillips T and Taylor  
16 K (2003) An overview of results from the Coupled Model Intercomparison Project.  
17 *Global and Planetary Change* 103-133.
- 18 Davis RE, Hayden BP, Gay DA, Phillips WL, Jones GV (1997) The North Atlantic  
19 subtropical anticyclone. *J Climate* 10:728–744.
- 20 Diem JE (2013) Comments on “Changes to the North Atlantic Subtropical High and Its  
21 Role in the Intensification of Summer Rainfall Variability in the Southeastern  
22 United States.” *J Climate* 26:679-682.
- 23 Enfield DB and Alfaro EJ (1999) The Dependence of Caribbean Rainfall on the  
24 Interaction of the Tropical Atlantic and Pacific Oceans. *J. Climate* 12:2093-2103.
- 25 Gamble DW, Parnell DB, and Curtis S (2008) Spatial variability of the Caribbean mid-  
26 summer drought and relation to north Atlantic high circulation. *Int J Climatol*  
27 28:343-350.

1 Giannini A, Kushnir Y, and Cane MA (2000) Interannual variability of Caribbean Rainfall,  
2 ENSO, and the Atlantic Ocean. *J Climate* 13:297-311.

3 Hawkins E and Sutton R (2011) The potential to narrow uncertainty in projections  
4 of regional precipitation change. *Clim Dyn* 37:407-418.

5 Houghton RW (1991) The relationship of Sea Surface Temperature to Thermocline Depth  
6 at Annual and Interannual Time Scales in the Tropical Atlantic Ocean. *J Geophys*  
7 *Res* 96:15,173-15,185.

8 Inoue M, Handoh IC, Bigg GR (2002) Bimodal distribution of tropical cyclogenesis in the  
9 Caribbean: characteristics and environmental factors. *J Clim* 15:2897-2905.

10 Jury MR (2009a) A quasi-decadal cycle in Caribbean climate. *J Geophys Res* 114,  
11 D13102, doi:10.1029/2009JD011741.

12 Jury MR (2009b) An intercomparison of observational, reanalysis, satellite, and coupled  
13 model data on mean rainfall in the Caribbean. *J. Hydrometeorology* 10:413-430.

14 Jury MR, Malmgren BA, and Winter A (2007) Subregional precipitation climate of the  
15 Caribbean and relationships with ENSO and NAO. *J Geophys Res* 112, D16107,  
16 doi:10.1029/2006JD007541.

17 Kelly P and Mapes B (2011) Zonal mean wind, the Indian monsoon, and July drying in the  
18 western Atlantic subtropics. *J Geophys Res* 116, D00Q07,  
19 doi:10.1029/2010JD015405.

20 Li L, Li W, and Deng Y (2013a) Summer rainfall variability over the Southeastern United  
21 States and its intensification in the 21st century as assessed by CMIP5 models, *J*  
22 *Geophys Res*, 118, 1-15, doi: 10.1002/jgrd.50136.

23 Li L, Li W, and Kushnir Y (2012) Variation of the North Atlantic subtropical high western  
24 ridge and its implication to Southeastern US summer precipitation. *Clim Dyn* 39,  
25 1401-1412.

1 Li W, Li L, Fu R, Deng Y, Wang H (2011) Changes to the North Subtropical High and Its  
2 Role in the Intensification of summer rainfall variability in the Southeastern United  
3 States. *J Clim* 24:1499-1506.

4 Li W, Li L, Fu R, Deng Y, and Wang H (2013b) Reply to “Comments on Changes to the  
5 North Atlantic Subtropical High and Its Role in the Intensification of Summer  
6 Rainfall Variability in the Southeastern United States.” *J. Climate*. 26:683-688.

7 Lienert F, Fyfe JC, and Merryfield WJ (2011) Do Climate Models Capture the Tropical  
8 Influence on North Pacific Sea Surface Temperature Variability? *J Clim* 24:6203-  
9 6209.

10 Magaña V, Amador JA, and Medina S (1999) The midsummer drought over Mexico and  
11 Central America. *J Clim* 12:1577-1588.

12 Malmgren BA., Winter A, and Chen D (1998) El Niño - Southern Oscillation and North  
13 Atlantic Oscillation Control of Climate in Puerto Rico. *J Clim* 11:2713-2717.

14 Martin ER and Schumacher C (2011) The Caribbean Low-Level Jet and its Relationship  
15 with Precipitation in IPCC AR4 Models. *J Clim* 24:5935-5950.

16 Meehl GA, Covey C, McAvaney B, Latif M, and Stouffer RJ (2005) Overview of the  
17 Coupled Model Intercomparison Project. *Bull Amer Meteor Soc* 86:89-93.

18 Moss RH and coauthors (2010) The next generation of scenarios for climate change  
19 research and assessment. *Nature* 463:747-756.

20 Munoz E, Busalacchi AJ, Nigam S, and Ruiz-Barradas A (2008) Winter and summer  
21 structure of the Caribbean low-level jet. *J Clim* 21:1260-1276.

22 Nakicenovic N and Swart R (Eds.) (2000) IPCC Special Report on Emissions Scenario.  
23 Cambridge University Press, UK.

24 Rahmstorf S, Cazenave A, Church J, Hansen J, Keeling R, Parker D, and Somerville R  
25 (2007) Recent Climate Observations Compared to Projections. *Science*, 316, 5825,  
26 709, DOI:10.1126/science.1136843.

1 Rauscher SA, Giorgi F, Diffenbaugh NS, and Seth A (2008) Extension and Intensification  
2 of the Meso-American mid-summer drought in the twenty-first century. *Clim Dyn*  
3 doi 10.1007/s00382-007-0357-1.

4 Reynolds RW, Rayner NA, Smith TM, Stokes DC, and Wang W (2002) An improved in  
5 situ and satellite SST analysis for climate. *J Clim* 15:1609-1625.

6 Small RJO, de Szoeke SP, Xie SP (2007) The Central American Midsummer Drought:  
7 Regional Aspects and Large-Scale Forcing. *J Clim* 20:4853-4873.

8 Stensrud DJ (1996) Importance of low-level jets to climate: A review. *J Clim* 9:1698-1711.

9 Stoner A, Hayhoe K, and Wuebbles D (2009) Assessing General Circulation Model  
10 Simulations of Atmospheric Teleconnection Patterns. *J Clim* 22:4348-4372.

11 Taylor KE, Stouffer RJ, and Meehl GA (2012) An Overview of CMIP5 and the  
12 experiment design. *Bull Am Meteorol Soc* 93:485-498.

13 Uppala SM and coauthors (2005) The ERA-40 re-analysis. *Q J R Meteorol Soc* 131:2961-  
14 3012.

15 Wang C (2007) Variability of the Caribbean Low-Level Jet and its relations to climate.  
16 *Clim Dyn* 29:411-422.

17 Wang C and Enfield DB (2003) A further study of the tropical Western Hemisphere warm  
18 pool. *J Clim* 16:1476-1493.

19 Wang C and Lee SK (2007) Atlantic warm pool, Caribbean low-level jet and their  
20 potential impact on atlantic hurricanes. *Geophys Res Lett* 34:L02703.  
21 doi:10.1029/2006GL028579.

22 Whyte F, Taylor M, Stephenson T, and Campbell J (2008) Features of the Caribbean low  
23 level jet. *J Climatology* 28:119-128.

24 Yu. L, Jin X, and Weller RA (2006) Role of Net Surface Heat Flux in Seasonal Variations  
25 of Sea Surface Temperature in the Tropical Atlantic Ocean. *J Clim* 19:6153-6169.

26

## 1 **Figure captions**

2

3 **Fig. 1** Location of stations over the Caribbean region ( $94^{\circ}\text{W} \sim 59^{\circ}\text{W}$ ,  $5^{\circ}\text{N} \sim 27^{\circ}\text{N}$ ) from  
4 GHCN and MIDAS (blue and light blue indicating selected and unselected stations,  
5 respectively) for precipitation. Three solid boxes indicate three sub-regions used to quantify  
6 surface climate variability: Region 1 (Central America), Region 2 (Central Caribbean,  
7 Antilles), and Region 3 (Eastern Caribbean, West Indies). The gridded lines indicate the  
8 horizontal resolutions of the GPCP fields, onto which the station data were objectively  
9 interpolated.

10 **Fig. 2** Climatological annual cycles of precipitation ( $\text{mm day}^{-1}$ ) averaged over the three  
11 Caribbean regions for the 30-year period from 1979 to 2008 based on (a) GPCP gridded data  
12 and (b) the merged and interpolated GHCN+MIDAS station data. A shaded area represents  
13  $\pm 1-\sigma$  (standard deviation) from the mean.

14 **Fig. 3** Scatter plots of the climatological means of sea surface temperature (SST) and  
15 precipitation averaged over each sub-regions for the 30-year period from 1979 to 2008 from  
16 observations (black closed circle), the CMIP3 (orange-red) and the CMIP5 (green-blue)  
17 simulations.

18 **Fig. 4** Normalized annual cycle of precipitation averaged over the regions, based on GPCP  
19 (solid black) and the CMIP3 simulations. The simulations are divided into three groups, one  
20 with a bimodal distribution, the second with a single peak with a Mid-Summer Drought, and  
21 the third one with a single peak without the MSD.

22 **Fig. 5** Same as **Fig. 4** except for CMIP5 simulations.

23 **Fig. 6** Monthly mean sea surface temperature (shaded), eddy geopotential height (black,  
24 contoured) at 925 hPa and zonal wind (blue, contoured) for each month of the Caribbean  
25 wet season from May to September. The contour interval of the zonal wind is  $2 \text{ m s}^{-1}$  and  
26 possible contours include -14, -12 and -10, while the eddy geopotential height is drawn with  
27 a contour interval of 20 m between 30 m and 110 m. The observation data (left) are based on  
28 NOAA SST from 1982 to 2008 and ERA-40 output from 1979 to 2001. The CMIP3  
29 simulations are used for composite calculation of three groups of models (Bimodal, Single  
30 with MSD, and Single).

31 **Fig. 7** Same as **Fig. 6** except for CMIP5 simulations.

32 **Fig. 8.** Latitude-Pressure cross-section of eddy geopotential height (shaded), zonal wind  
33 (blue, contoured) and meridional and vertical wind (vector) averaged over longitude  
34 between  $60^{\circ}\text{W}$  and  $80^{\circ}\text{W}$ . The contour interval of the zonal wind is  $2 \text{ m s}^{-1}$  and possible

1 contours include -14, -12, -10, -8, and -6. CMIP5 simulations are used for composite  
2 calculation of three groups of models (Bimodal, Single with MSD, and Single).  
3 **Fig. 9** Monthly mean of the vertically integrated moisture flux convergence/divergence  
4 (shaded) and the horizontal wind at 925 hPa (arrow). CMIP5 simulations are used for  
5 composite calculation of three groups of models (Bimodal, Single with MSD, and Single).

6

7

8

9

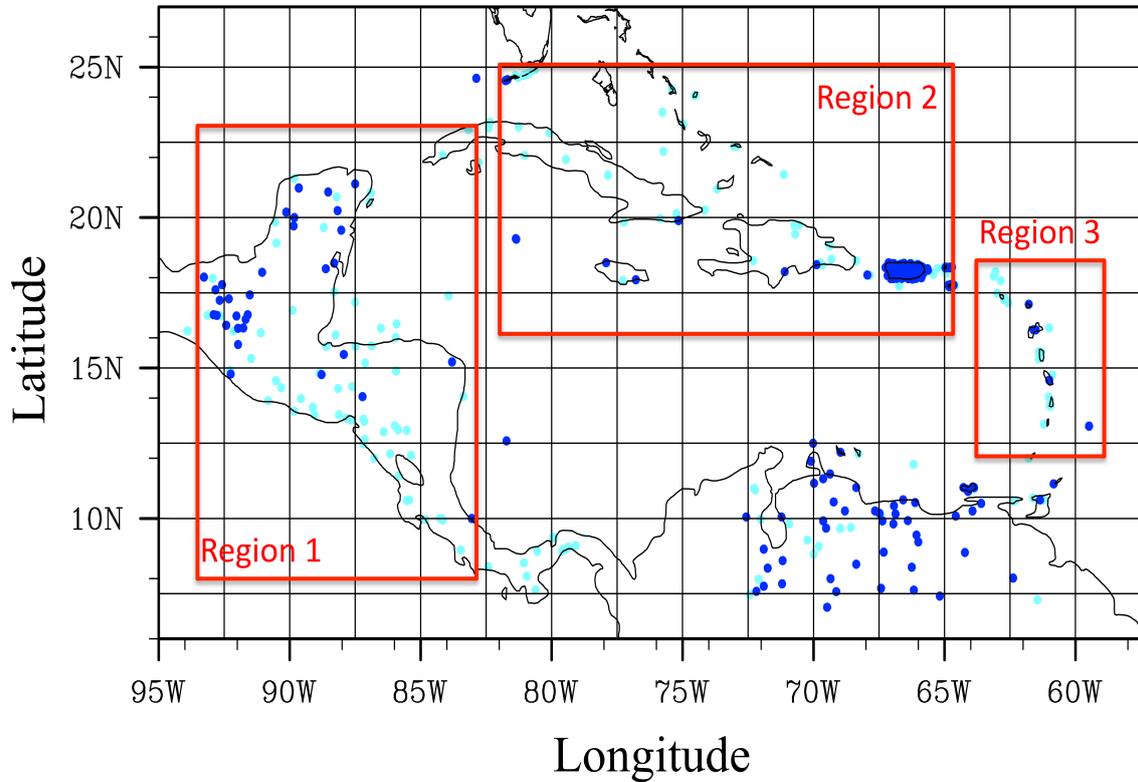
10

11

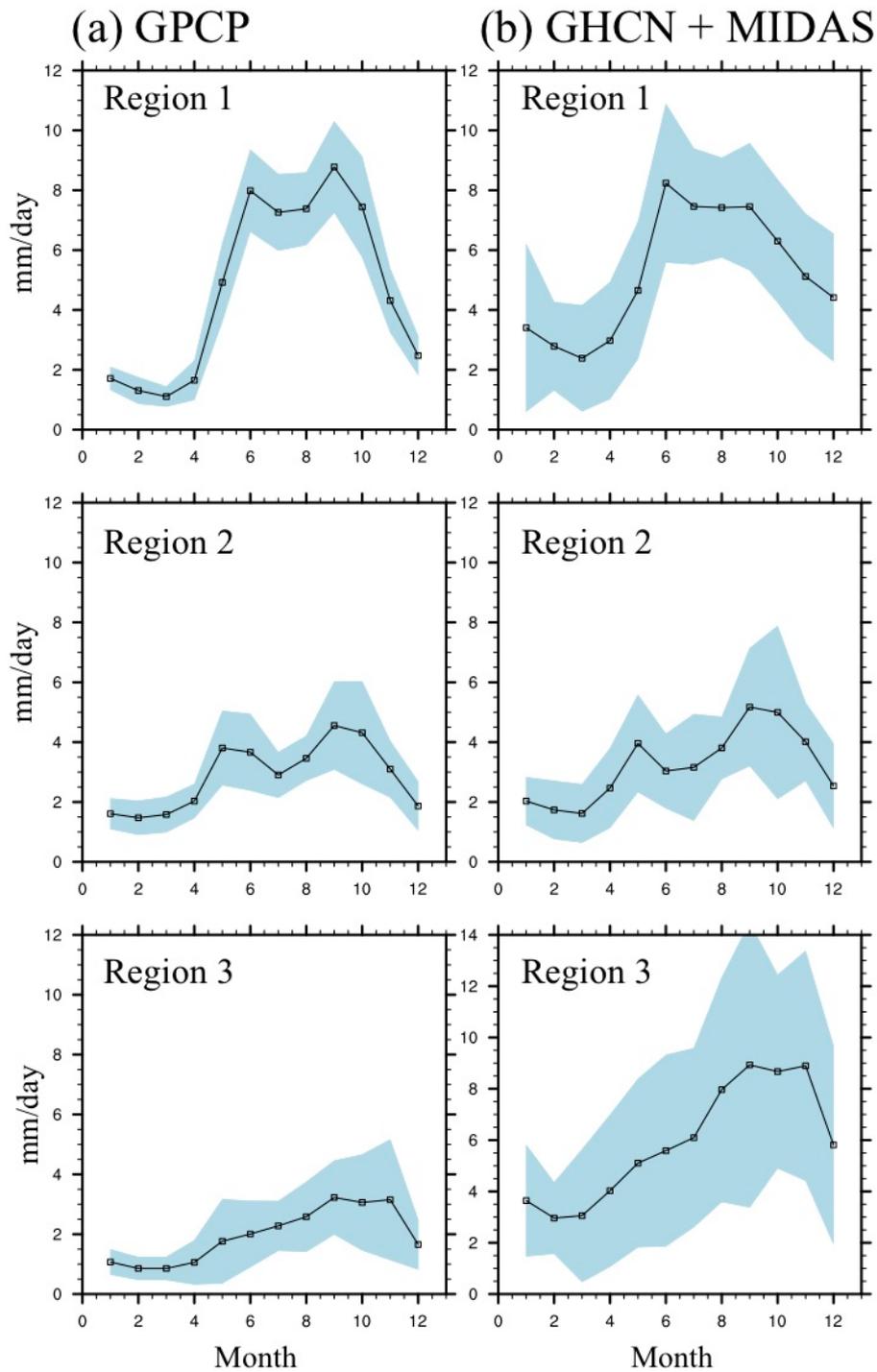
## 12 **Table captions**

13

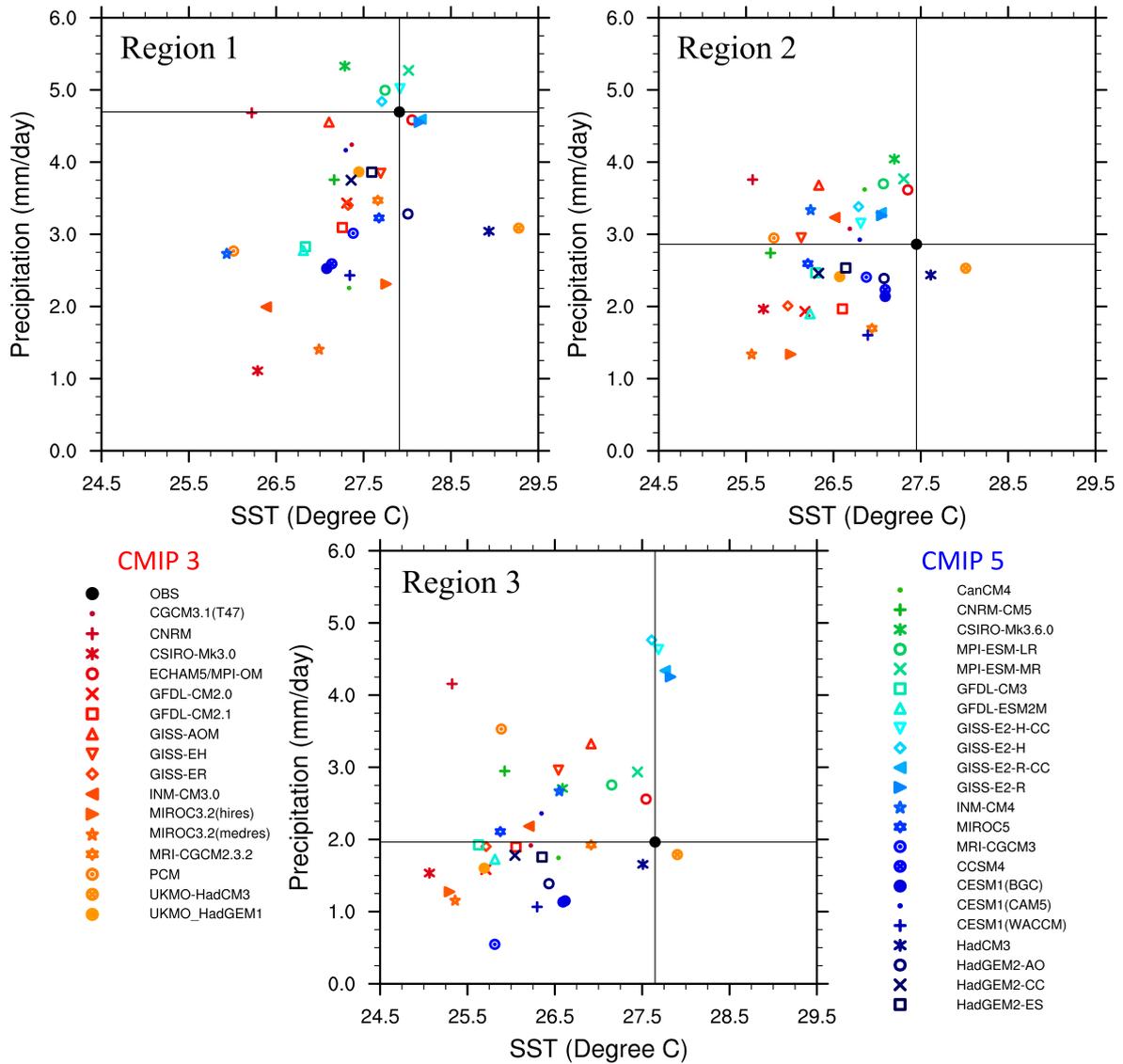
14 **Table 1** Primary descriptions of the CMIP3 and CMIP5 models used in this analysis,  
15 including their names, provenance, and horizontal and vertical resolution. The three  
16 different indices (“B”, “SM”, and “S”), posted next to the model names, represent  
17 “Bimodal”, “Single w/MSD”, and “Single”, respectively. These are the three categories the  
18 models were divided into, based on their ability to simulate the annual cycle of precipitation  
19 over the Caribbean.



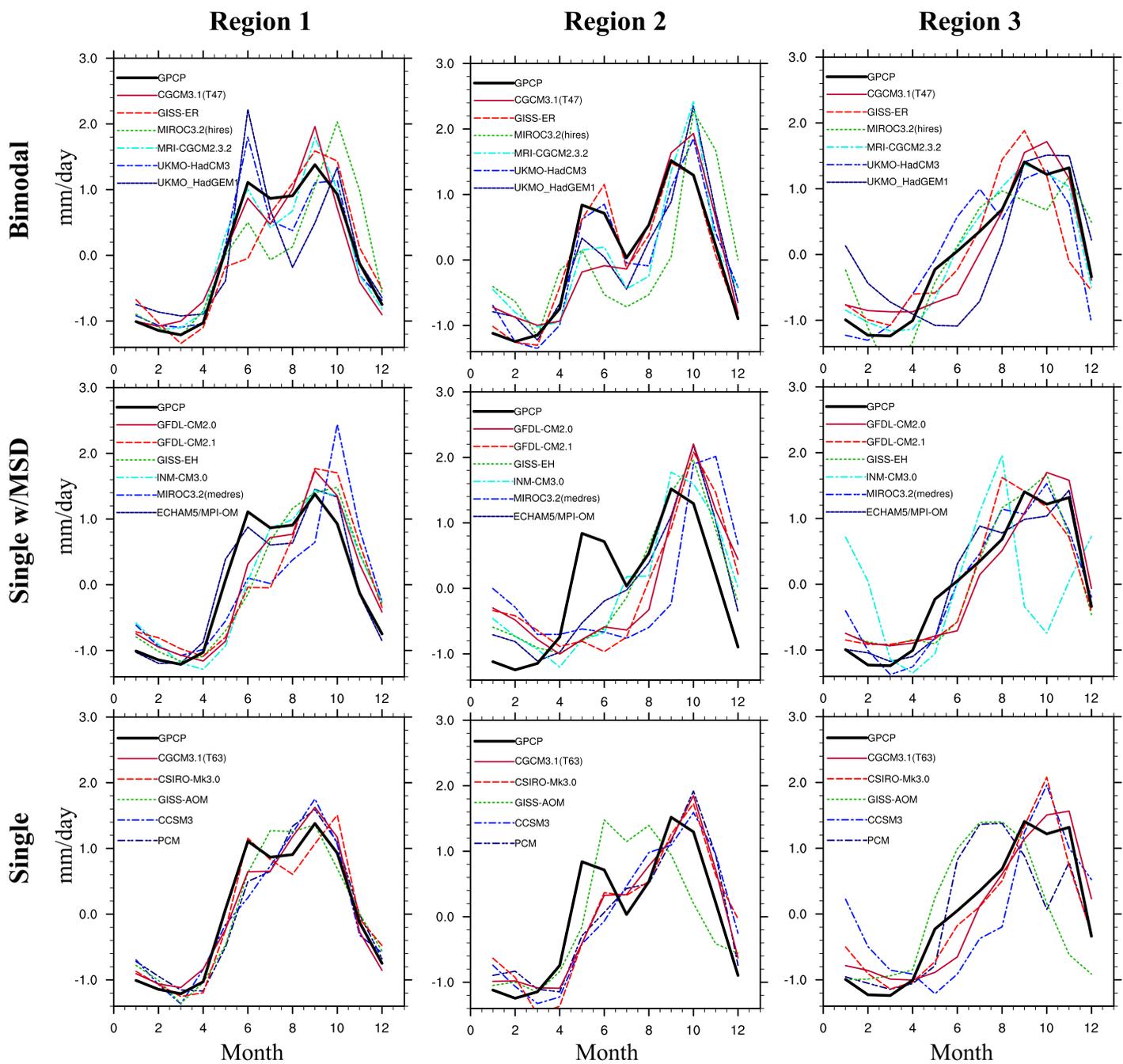
**Fig. 1** Location of stations over the Caribbean region ( $94^{\circ}\text{W} \sim 59^{\circ}\text{W}$ ,  $5^{\circ}\text{N} \sim 27^{\circ}\text{N}$ ) from GHCN and MIDAS (blue and light blue indicating selected and unselected stations, respectively) for precipitation. Three solid boxes indicate three sub-regions used to quantify surface climate variability: Region 1 (Central America), Region 2 (Central Caribbean, Antilles), and Region 3 (Eastern Caribbean, West Indies). The gridded lines indicate the horizontal resolutions of the GPCP fields, onto which the station data were objectively interpolated.



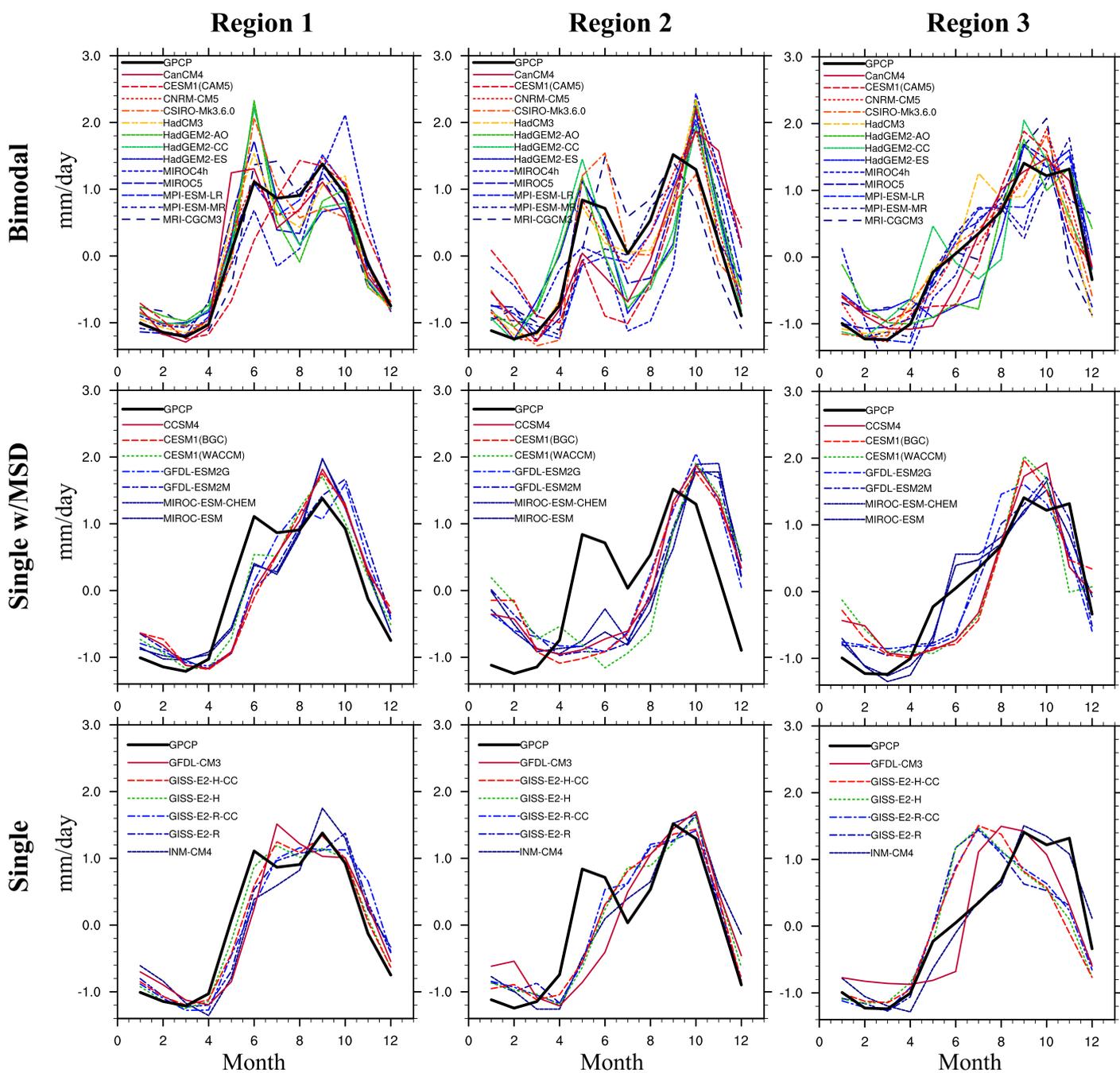
**Fig. 2** Climatological annual cycles of precipitation ( $\text{mm day}^{-1}$ ) averaged over the three Caribbean regions for the 30-year period from 1979 to 2008 based on (a) GPCP grid data and (b) the merged and interpolated GHCN+MIDAS station data. A shaded area represents  $\pm 1\text{-}\sigma$  (standard deviation) from the mean.



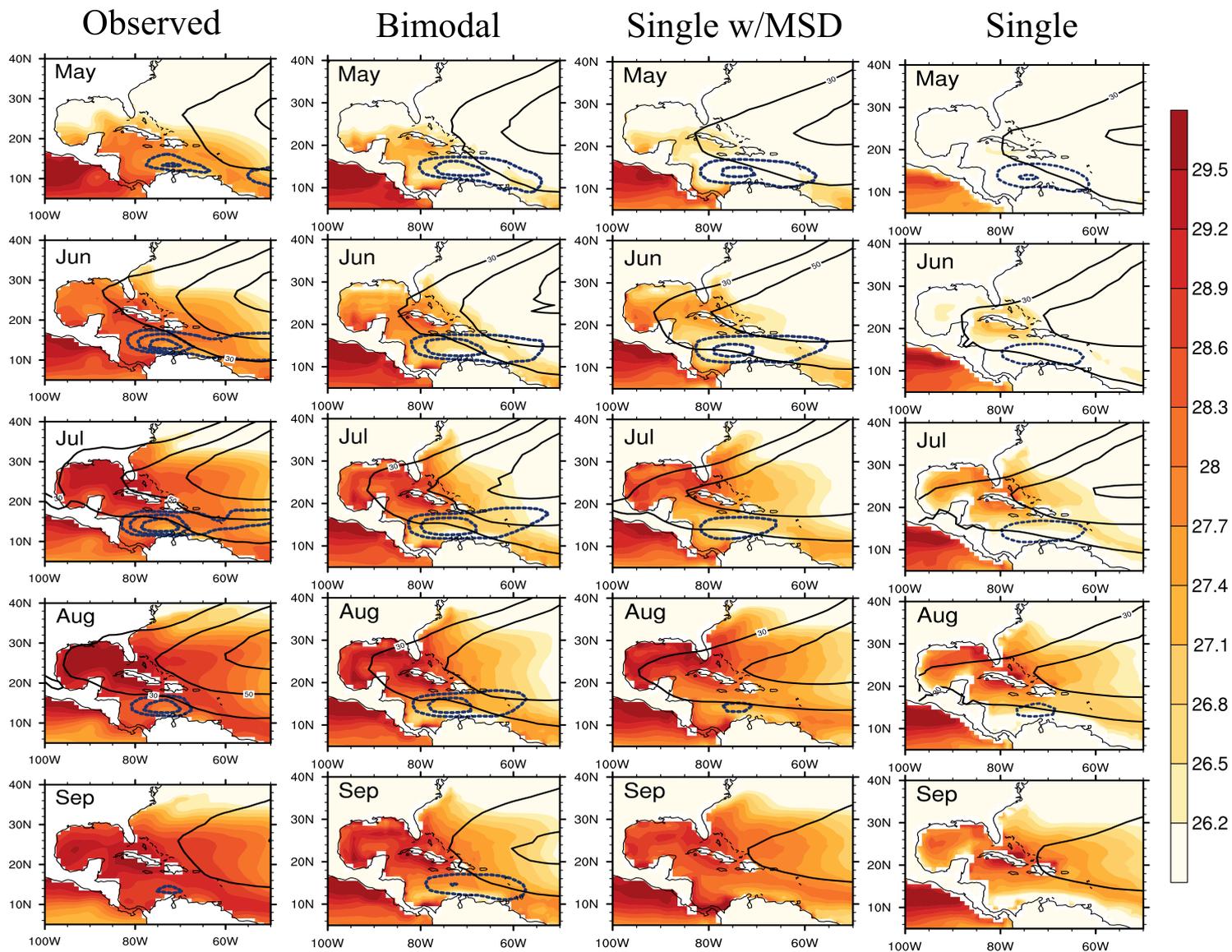
**Fig. 3** Scatter plots of the climatological means of sea surface temperature (SST) and precipitation averaged over each sub-regions for the 30-year period from 1979 to 2008 from observations (black closed circle), the CMIP3 (orange-red) and the CMIP5 (green-blue) simulations.



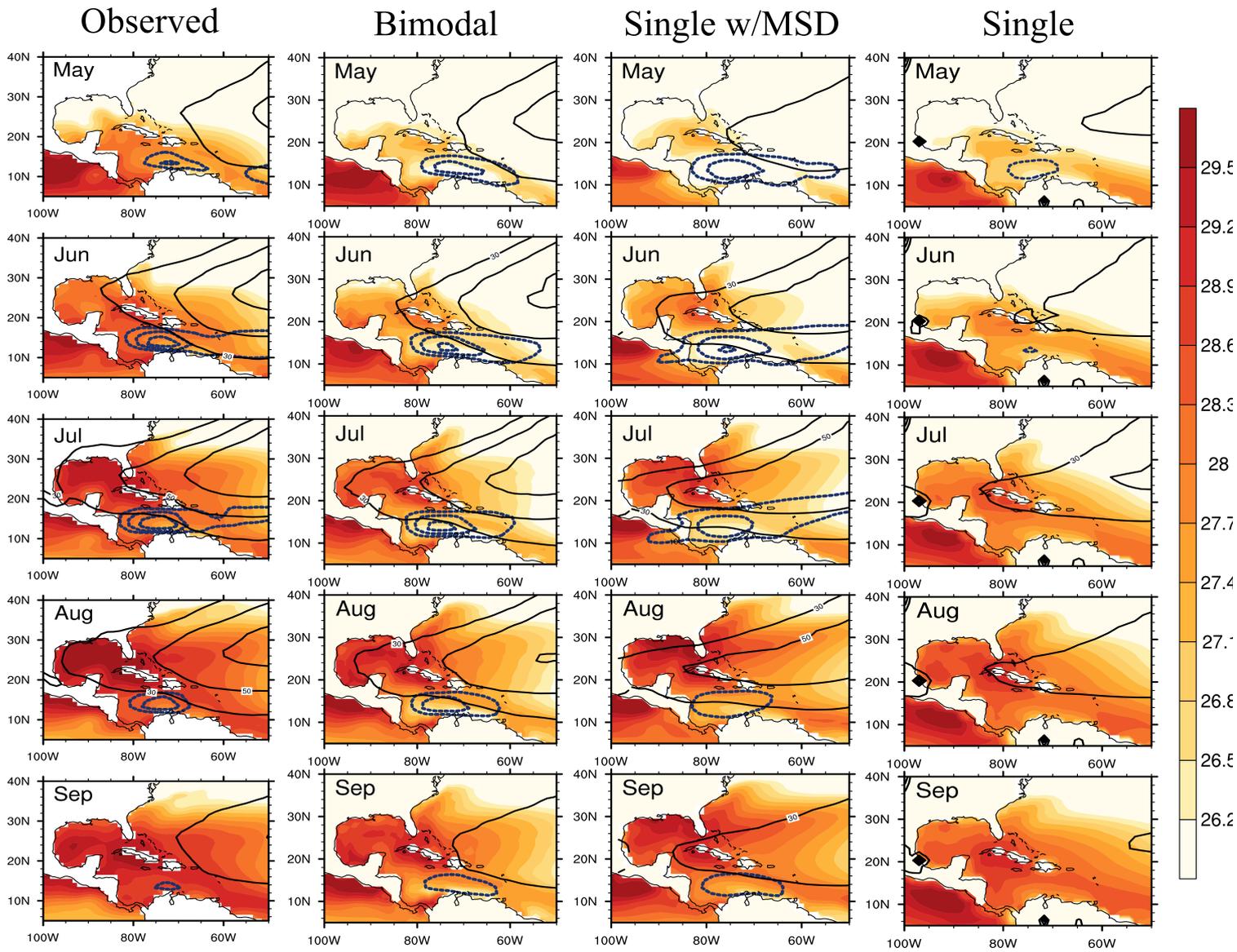
**Fig. 4** Normalized annual cycle of precipitation averaged over the regions, based on GPCP (solid black) and the CMIP3 simulations. The simulations are divided into three groups, one with a bimodal distribution, the second with a single peak with a Mid-Summer Drought, and the third one with a single peak without the MSD.



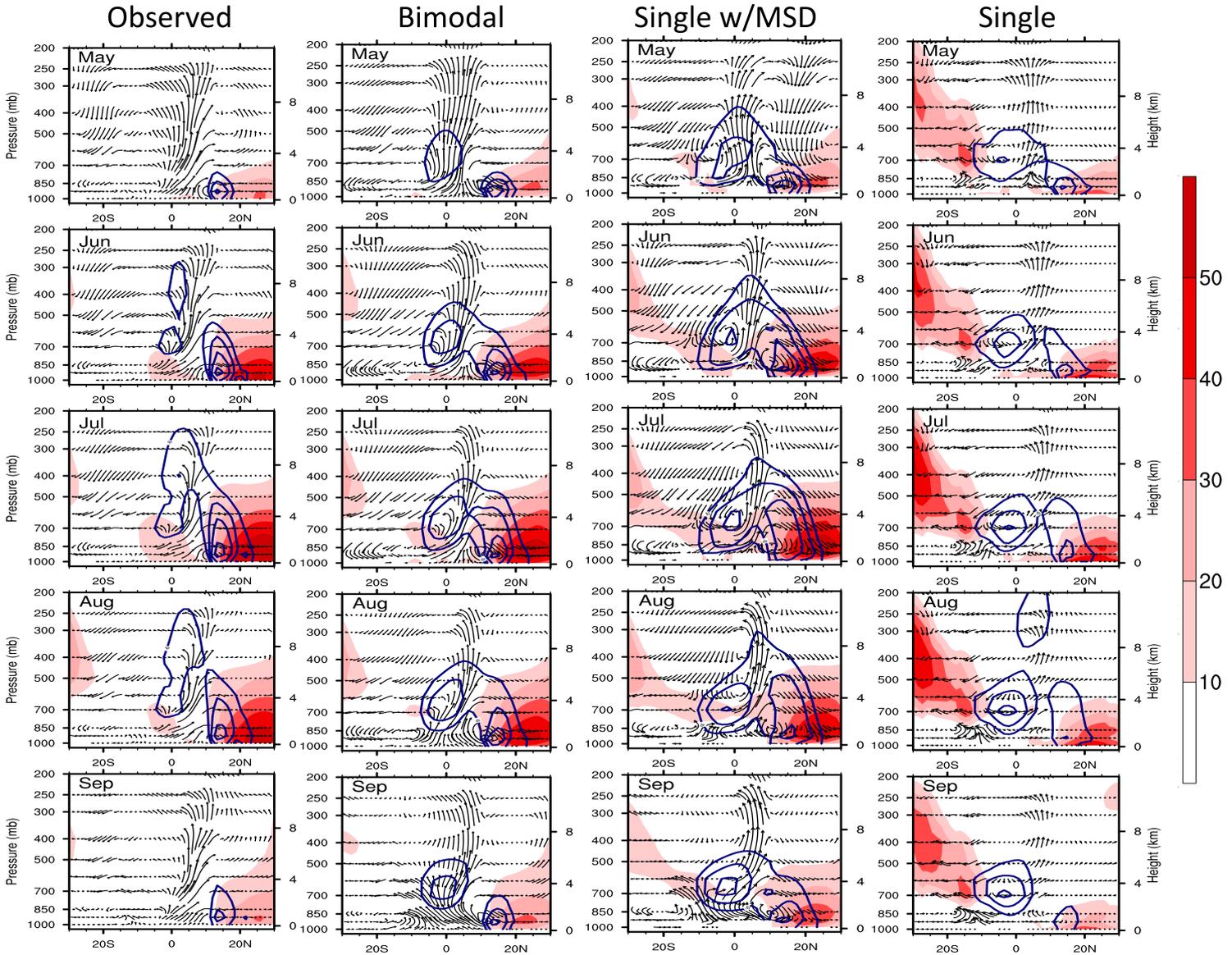
**Fig. 5** Same as **Fig. 4** except for CMIP5 simulations.



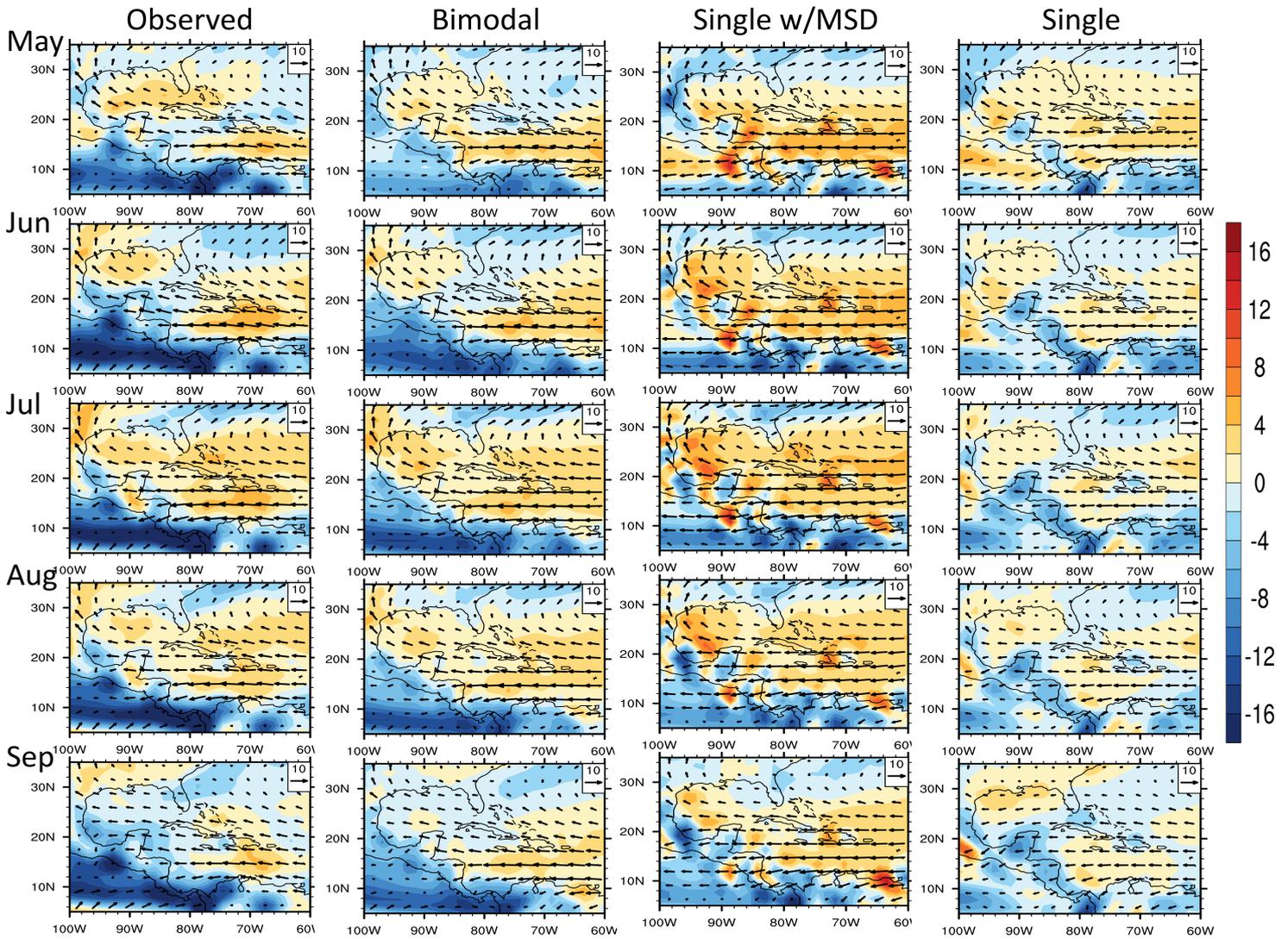
**Fig. 6** Monthly mean sea surface temperature (shaded), eddy geopotential height (black, contoured) at 925 hPa and zonal wind (blue, contoured) for each month of the Caribbean wet season from May to September. The contour interval of the zonal wind is  $2 \text{ m s}^{-1}$  and possible contours include -14, -12 and -10, while the eddy geopotential height is drawn with a contour interval of 20 m between 30 m and 110 m. The observation data (left) are based on NOAA SST from 1982 to 2008 and ERA-40 output from 1979 to 2001. The CMIP3 simulations are used for composite calculation of three groups of models (Bimodal, Single with MSD, and Single).



**Fig. 7** Same as **Fig. 6** except for CMIP5 simulations.



**Fig. 8.** Latitude-Pressure cross-section of eddy geopotential height (shaded), zonal wind (blue, contoured) and meridional and vertical wind (vector) averaged over longitude between  $60^{\circ}\text{W}$  and  $80^{\circ}\text{W}$ . The contour interval of the zonal wind is  $2 \text{ m s}^{-1}$  and possible contours include -14, -12, -10, -8, and -6. CMIP5 simulations are used for composite calculation of three groups of models (Bimodal, Single with MSD, and Single).



**Fig. 9** Monthly mean of the vertically integrated moisture flux convergence/divergence (shaded) and the horizontal wind at 925 hPa (arrow). CMIP5 simulations are used for composite calculation of three groups of models (Bimodal, Single with MSD, and Single).

Modeling center	CMIP3	CMIP5	Resolution (Atmosphere)	Variables						
				sst	zg	u	v	w	q	
Canadian Centre for Climate Modelling and Analysis, Canada	CGCM3.1(T47)	<b>B</b>	T47 (2.8°) L31	✓	✓	✓				
	CGCM3.1(T63)	<b>S</b>	T63 (1.9°) L31							
Center for Climate System Research (University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan		CanCM4	<b>B</b>	T63 (1.9°) L35	✓	✓	✓	✓		
	MIROC3.2 (hires)	<b>B</b>	T106 (1.1°) L56	✓	✓	✓				
	MIROC3.2 (medres)	<b>SM</b>	T42 (2.8°) L20	✓	✓	✓				
		MIROC4h	<b>B</b>	T213(0.6°) L56 w/aerosol		✓	✓	✓	✓	✓
		MIROC5	<b>B</b>	T85(1.4°) L40	✓	✓	✓	✓		✓
		MIROC-ESM	<b>SM</b>	T42 (2.8°hh) L80 w/ carbon cycle		✓	✓	✓	✓	✓
Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	CNRM-CM3	<i>N/A</i>	T63 (1.9° x 1.9°) L45	✓						
		CNRM-CM5	<b>B</b>	T127(1.4°) L91	✓	✓	✓	✓	✓	✓
CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	CSIRO-Mk3.0	<b>S</b>	T63 (1.9° x 1.9°) L18	✓	✓	✓				
		CSIRO-Mk3.6.0	<b>B</b>		✓	✓	✓	✓	✓	✓
Geophysical Fluid Dynamics Laboratory (GFDL), USA	GFDL-CM2.0	<b>SM</b>		✓	✓	✓				
	GFDL-CM2.1	<b>SM</b>		✓	✓	✓				
		GFDL-CM3	<b>S</b>	2.5° x 2.0° L24	✓	✓	✓	✓	✓	✓
		GFDL-ESM2G	<b>SM</b>			✓	✓	✓	✓	✓
Hadley Centre for Climate Prediction and Research/Met Office, UK		GFDL-ESM2M	<b>SM</b>		✓	✓	✓	✓	✓	✓
	UKMO-HadCM3	<b>B</b>	2.5° x 3.75° L19	✓	✓	✓				
		HadCM3	<b>B</b>		✓	✓	✓	✓	✓	
	UKMO-HadGEM1	<b>B</b>		✓	✓					
		HadGEM2-AO	<b>B</b>	1.25° x 1.875° L38	✓	✓	✓	✓	✓	
		HadGEM2-CC	<b>B</b>		✓	✓				
Institute for Numerical Mathematics, Russia		HadGEM2-ES	<b>B</b>		✓	✓	✓	✓	✓	✓
	INM-CM3.0	<b>SM</b>	4° x 5° L21	✓		✓				
Max Planck Institute for Meteorology, Germany		INM-CM4	<b>S</b>	1.5° x 2° L21	✓	✓	✓	✓	✓	✓
	ECHAM5/MPI-OM	<b>SM</b>	T63 (1.9°) L31	✓	✓	✓				
		MPI-ESM-LR	<b>B</b>	T63 (1.9°) L47	✓					
Meteorological Research Institute, Japan		MPI-ESM-MR	<b>B</b>	T63 (1.9°) L95	✓	✓	✓	✓	✓	✓
	MRI-CGCM2.3.2	<b>B</b>	T42 (2.8°) L30	✓	✓	✓				
		MRI-CGCM3	<b>B</b>	T42 (2.8°) L46	✓	✓	✓	✓	✓	✓
National Aeronautics and Space Administration (NASA)/ Goddard Institute for Space Studies (GISS), USA	GISS-AOM	<b>S</b>	3° x 4° L12	✓	✓	✓				
	GISS-EH	<b>SM</b>	4° x 5° L20	✓	✓					
	GISS-ER	<b>B</b>		✓	✓					
		GISS-E2-H	<b>S</b>		✓	✓	✓	✓	✓	✓
		GISS-E2-H-CC	<b>S</b>	2° x 2.5° L40	✓	✓	✓	✓	✓	✓
		GISS-E2-R	<b>S</b>		✓	✓	✓	✓	✓	✓
National Center for Atmospheric Research, USA		GISS-E2-R-CC	<b>S</b>		✓	✓	✓	✓	✓	✓
	CCSM3	<b>S</b>	T85 (1.4°) L26		✓	✓				
	PCM	<b>S</b>	T42 (2.8°) L26	✓	✓	✓				
National Science Foundation, Department of Energy, National Center for Atmospheric Research		CCSM4	<b>SM</b>	0.9° x 1.25° L26	✓	✓	✓	✓	✓	✓
		CESM1(BGC)	<b>SM</b>		✓	✓	✓	✓	✓	✓
		CESM1(CAM5)	<b>B</b>	1° x 1° L31	✓	✓	✓	✓	✓	✓
	CESM1(WACCM)	<b>SM</b>	2° x 2° L31	✓						

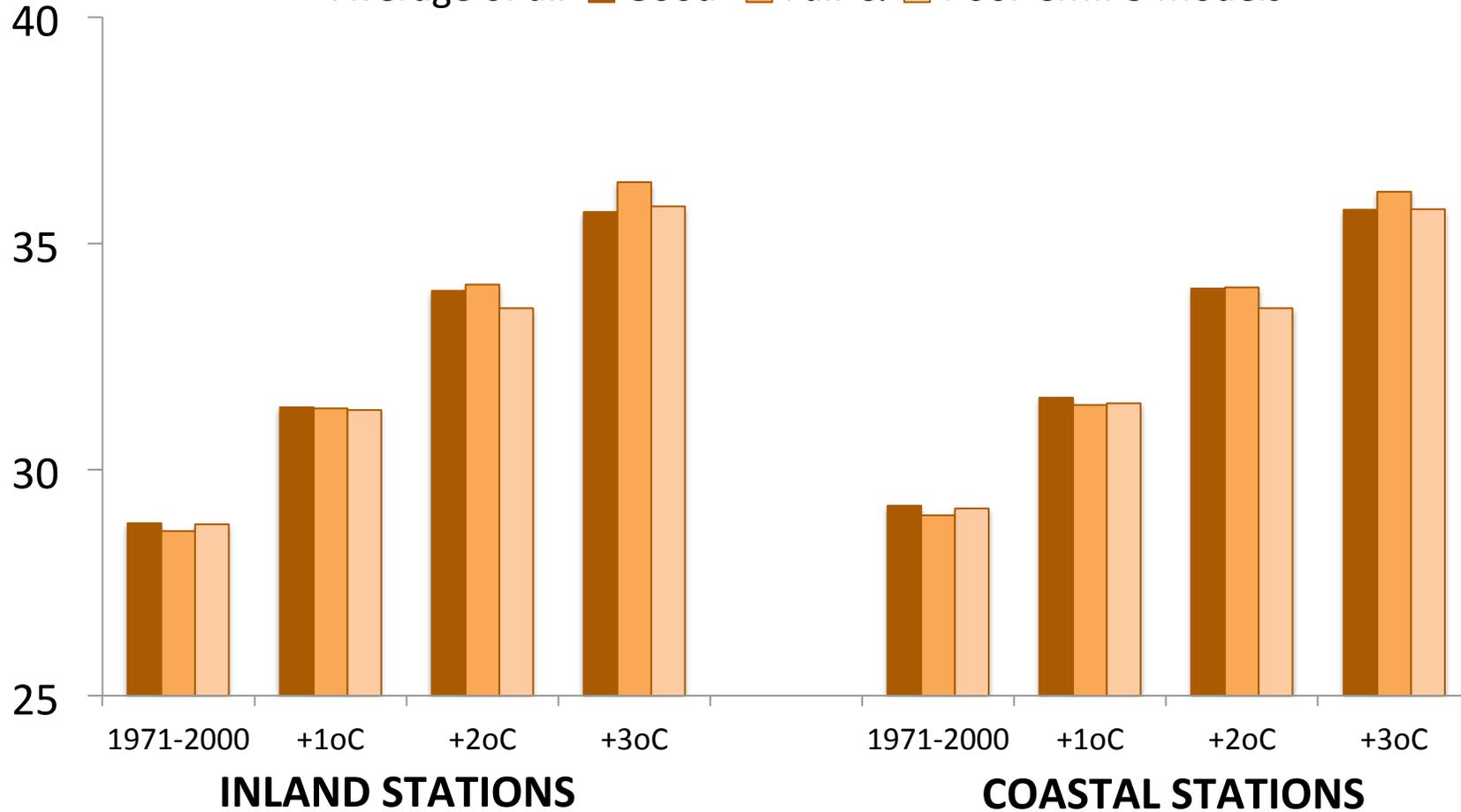
**Table 1** Primary descriptions of the CMIP3 and CMIP5 models used in this analysis, including their names, provenance, and horizontal and vertical resolution. The three different indices (“B”, “SM”, and “S”), posted next to the model names, represent “Bimodal”, “Single w/MSD”, and “Single”, respectively. These are the three categories the models were divided into, based on their ability to simulate the annual cycle of precipitation over the Caribbean.

## **APPENDIX C. Climate Projections for Puerto Rico: Secondary Indicator Plots**

The full list of 85 climate indicators is provided in Table 2 of the report.

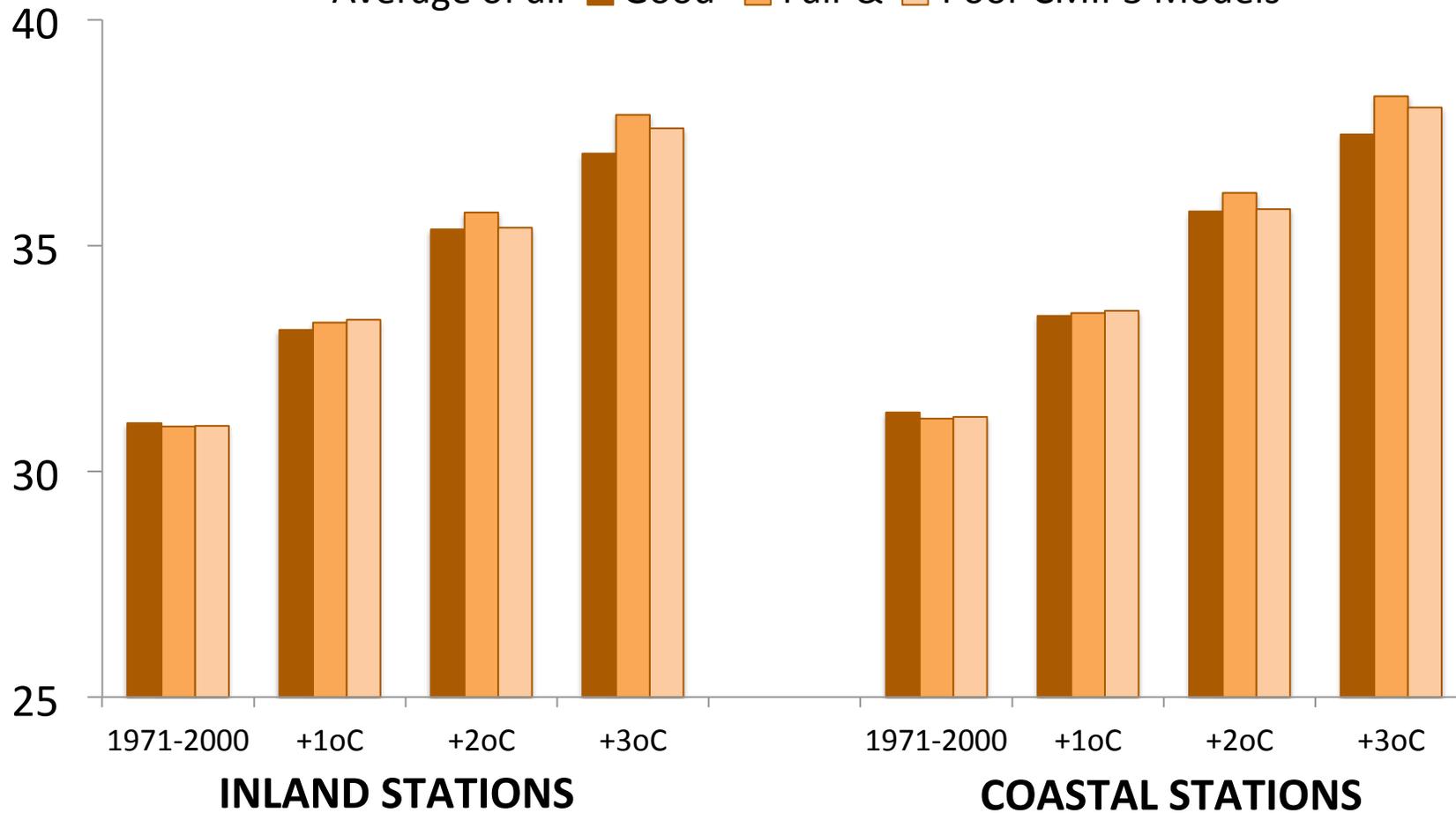
# Mean Daily Maximum Temperature - Dry Season

Average of all **Good** **Fair &** **Poor CMIP3 Models**



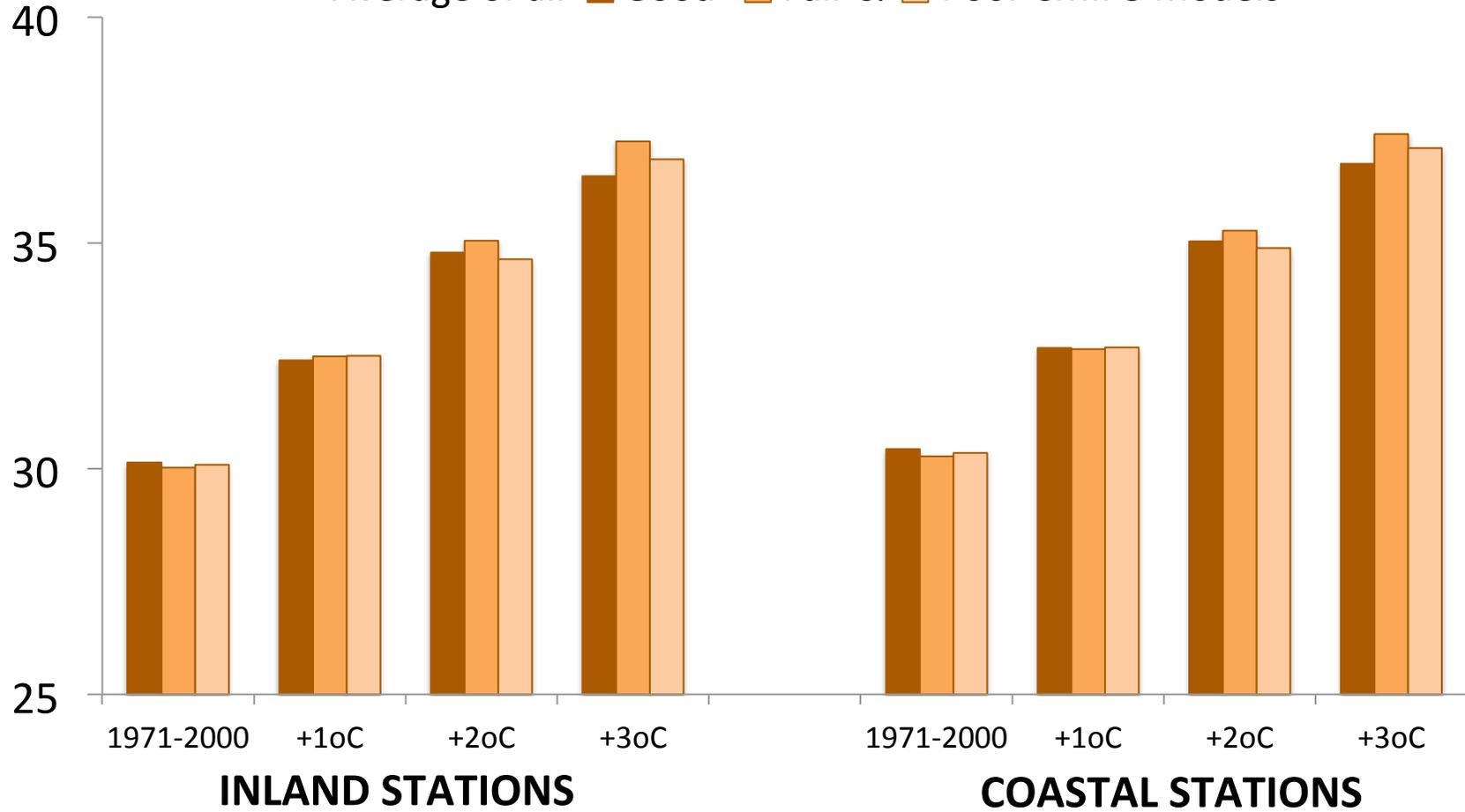
# Mean Daily Maximum Temperature - Wet Season

Average of all **Good** **Fair &** **Poor CMIP3 Models**



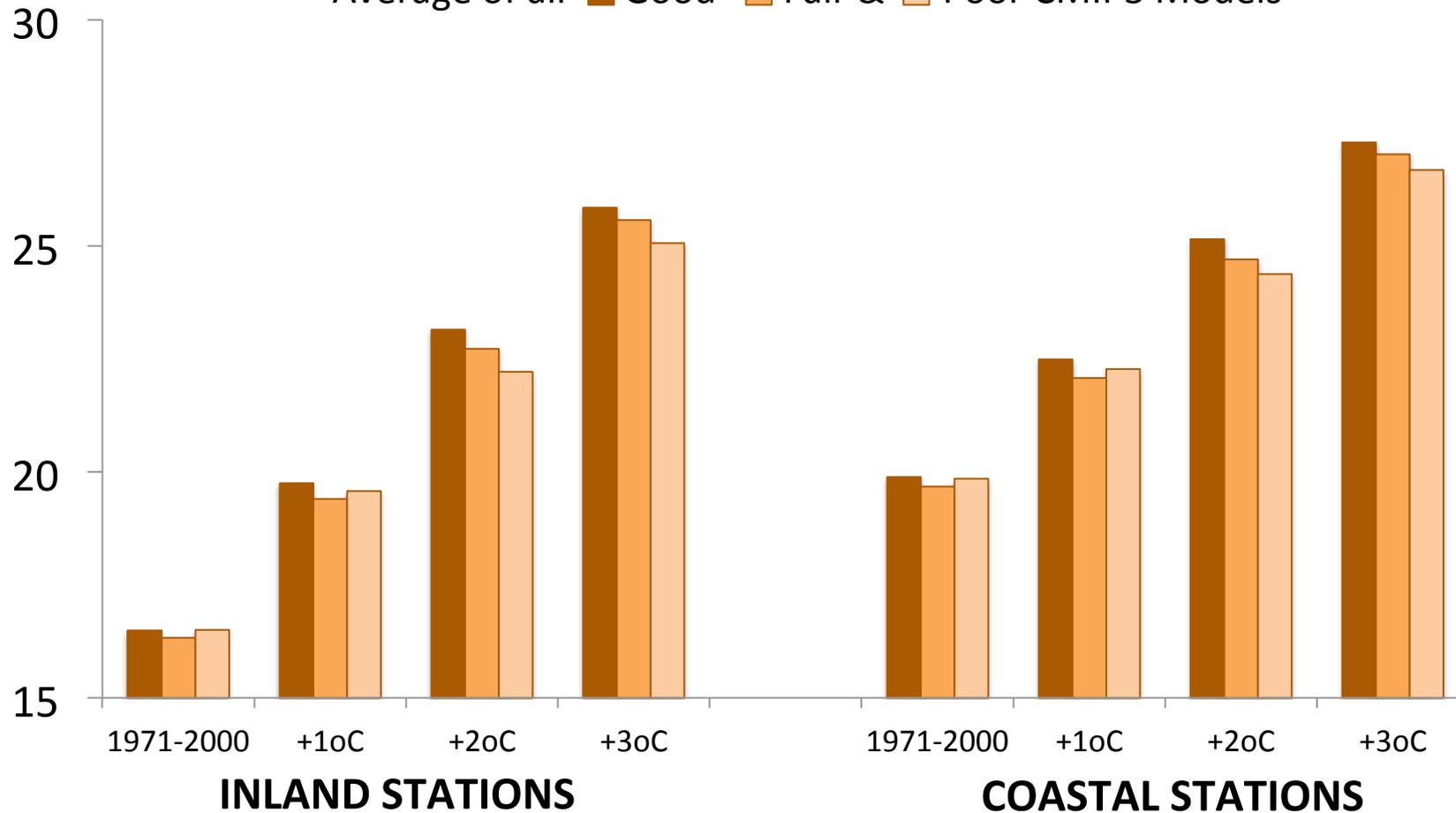
# Mean Daily Maximum Temperature - Annual

Average of all **Good** **Fair &** **Poor CMIP3 Models**

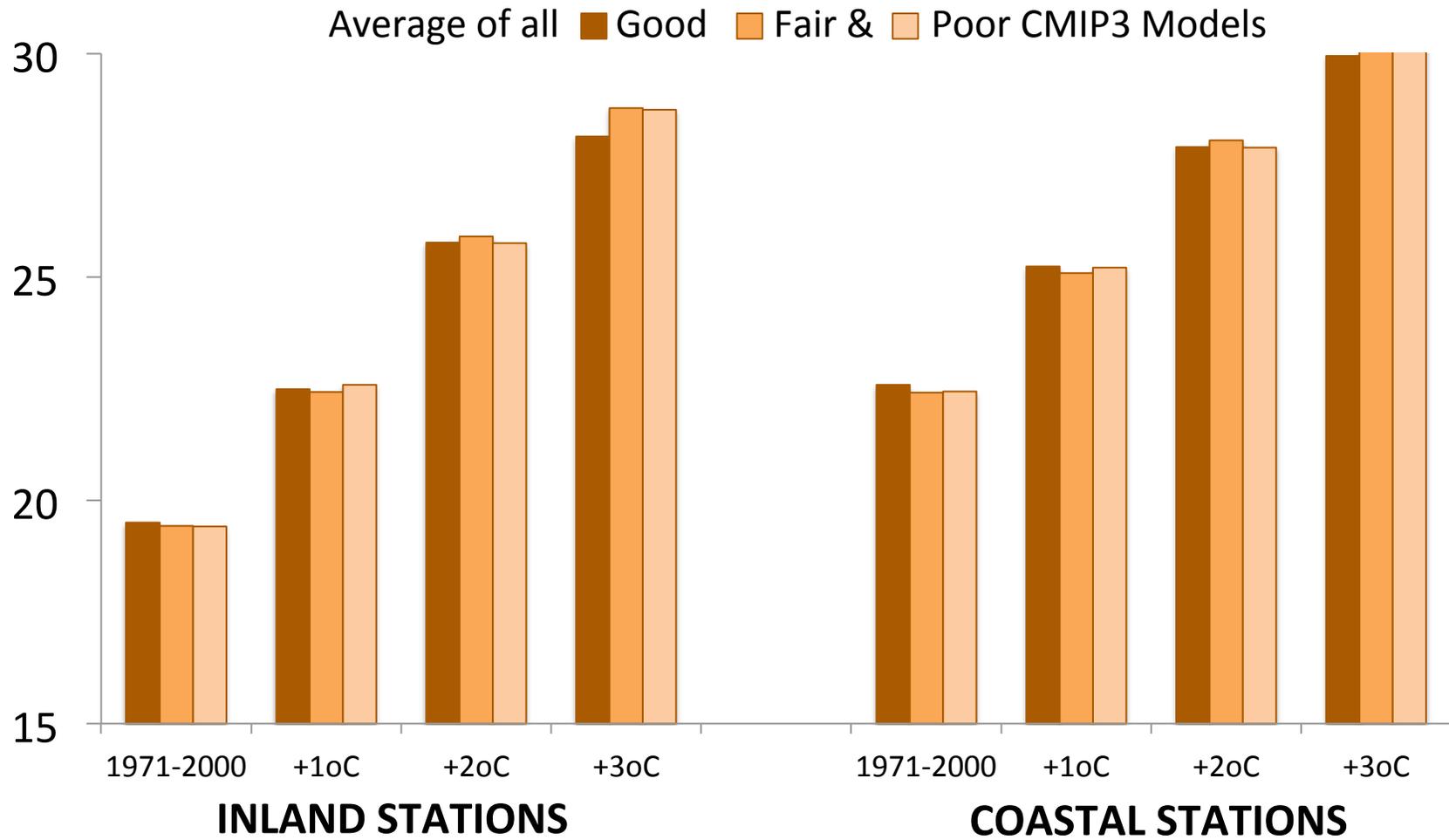


# Mean Nighttime Minimum Temperature - Dry Season

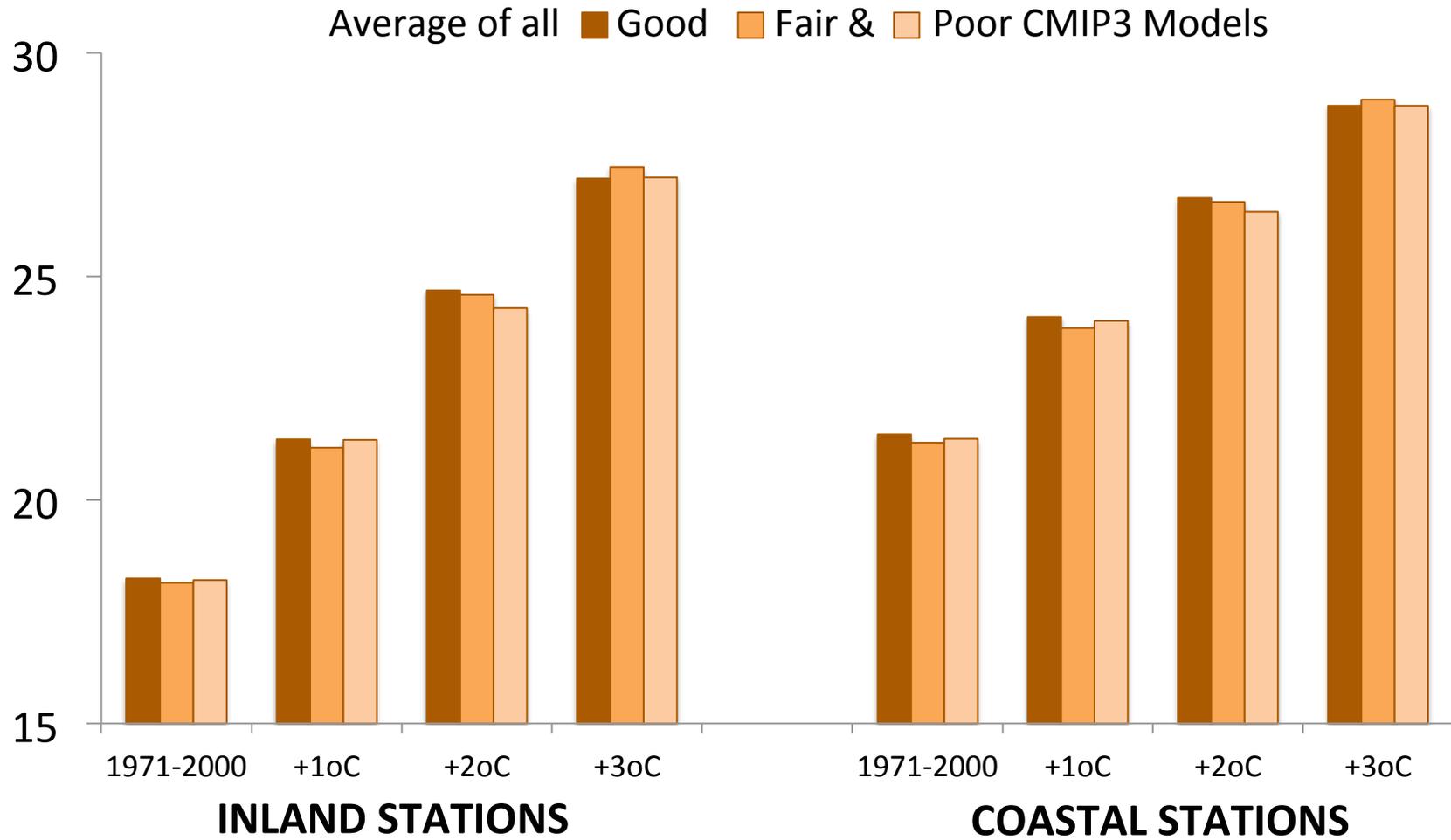
Average of all **Good** **Fair &** **Poor CMIP3 Models**



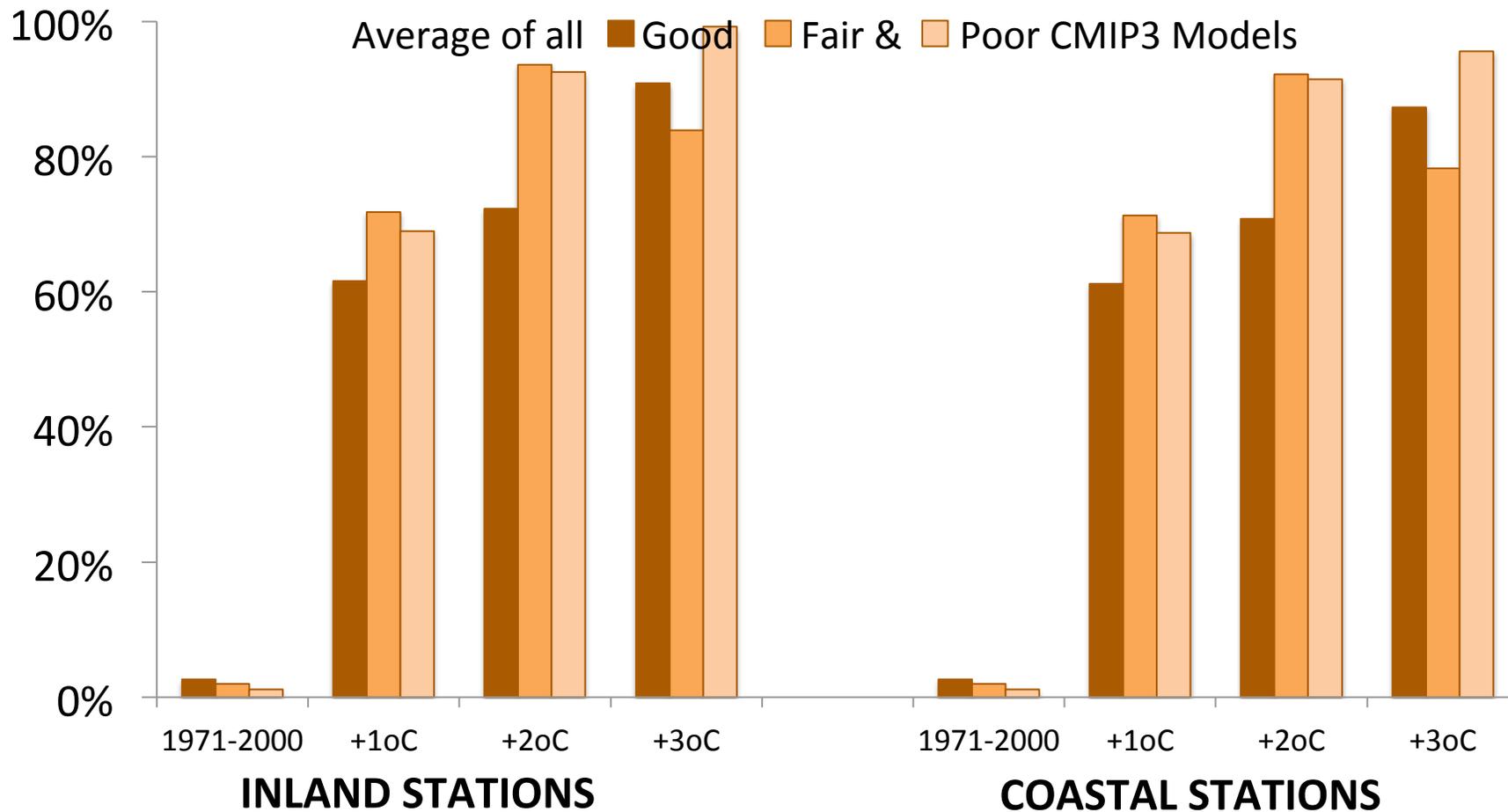
# Mean Nighttime Minimum Temperature - Wet Season



# Mean Nighttime Minimum Temperature - Annual

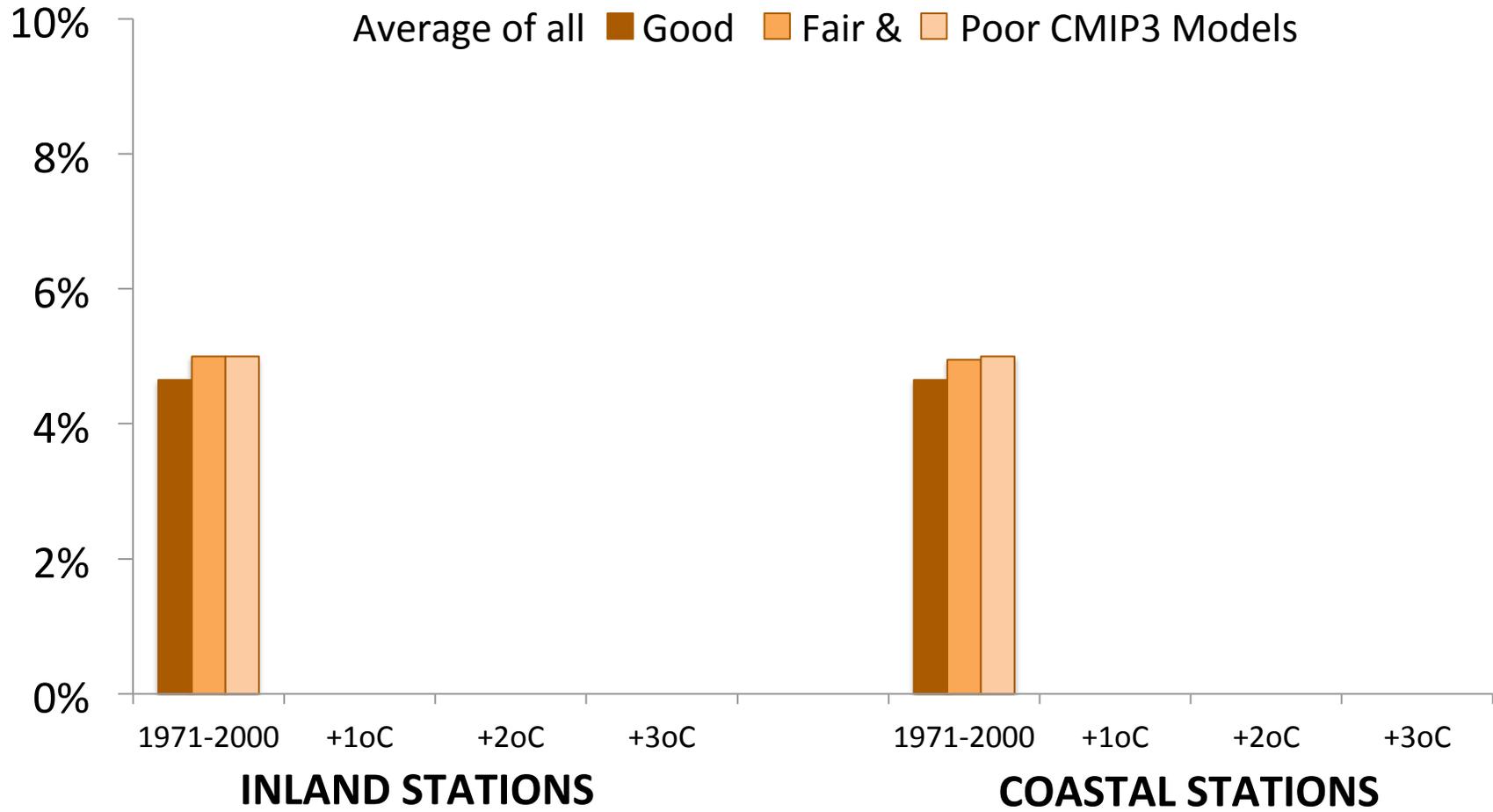


# Percentage of wet seasons warmer than hottest historical season

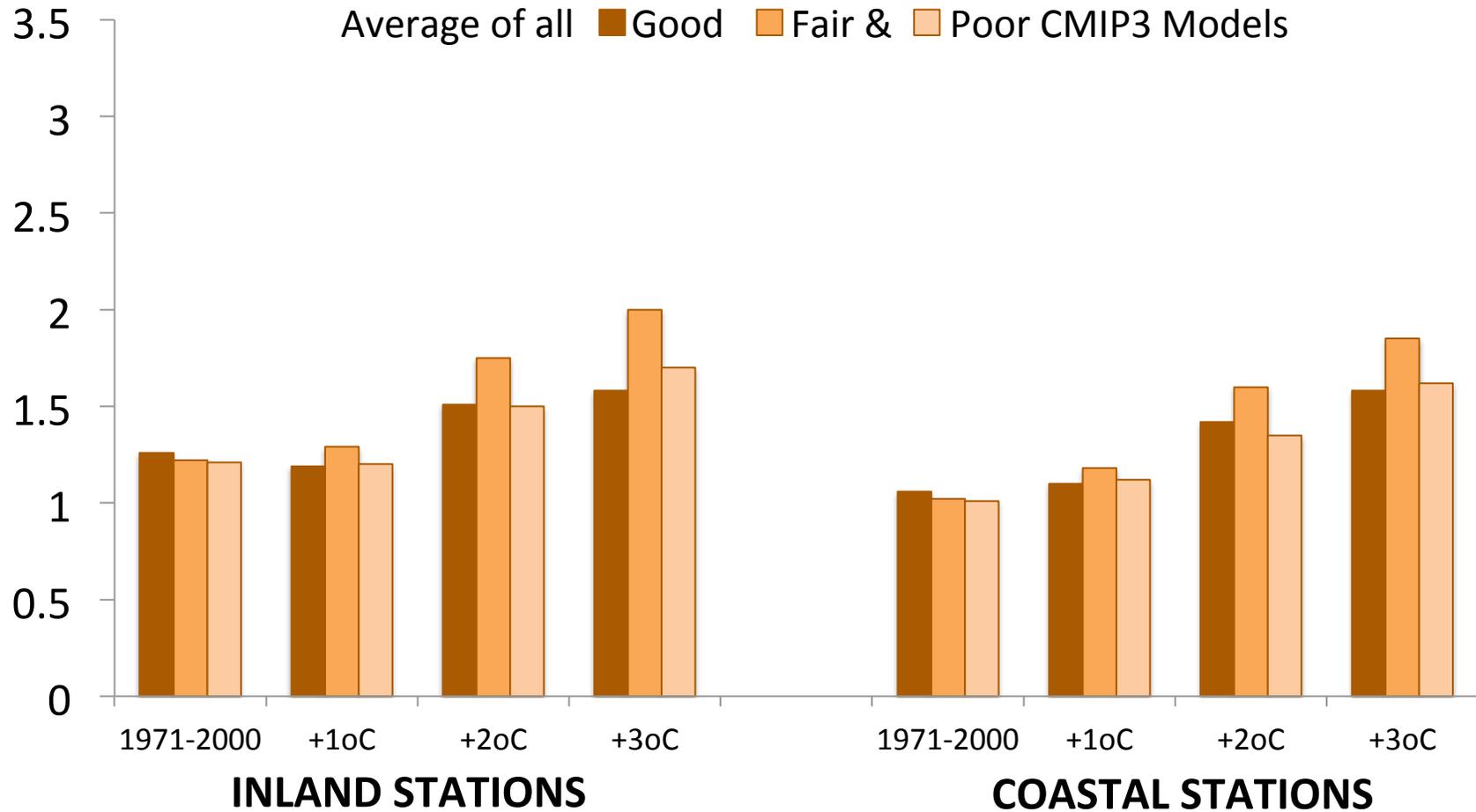


# Percentage of years colder than coldest historical year

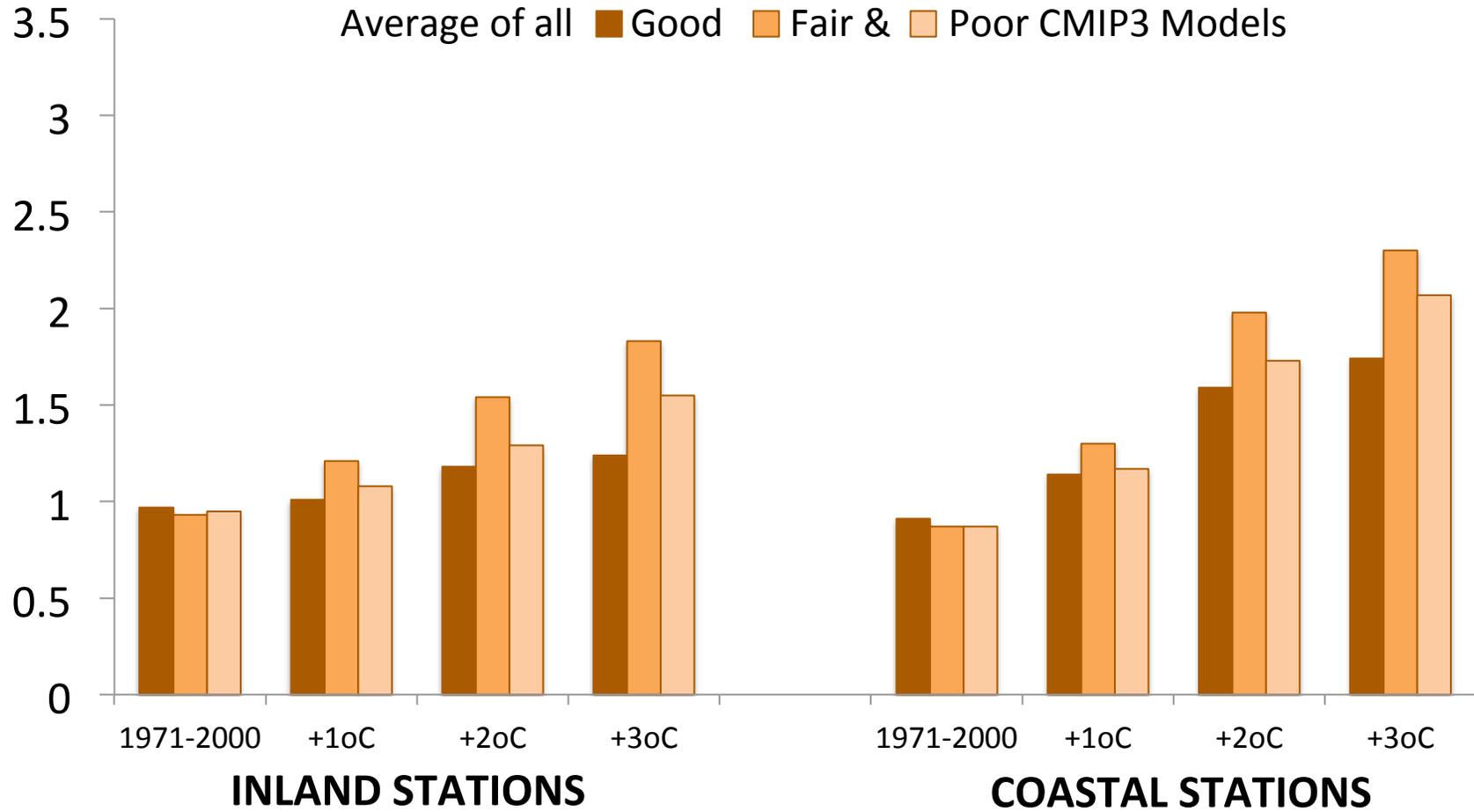
Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



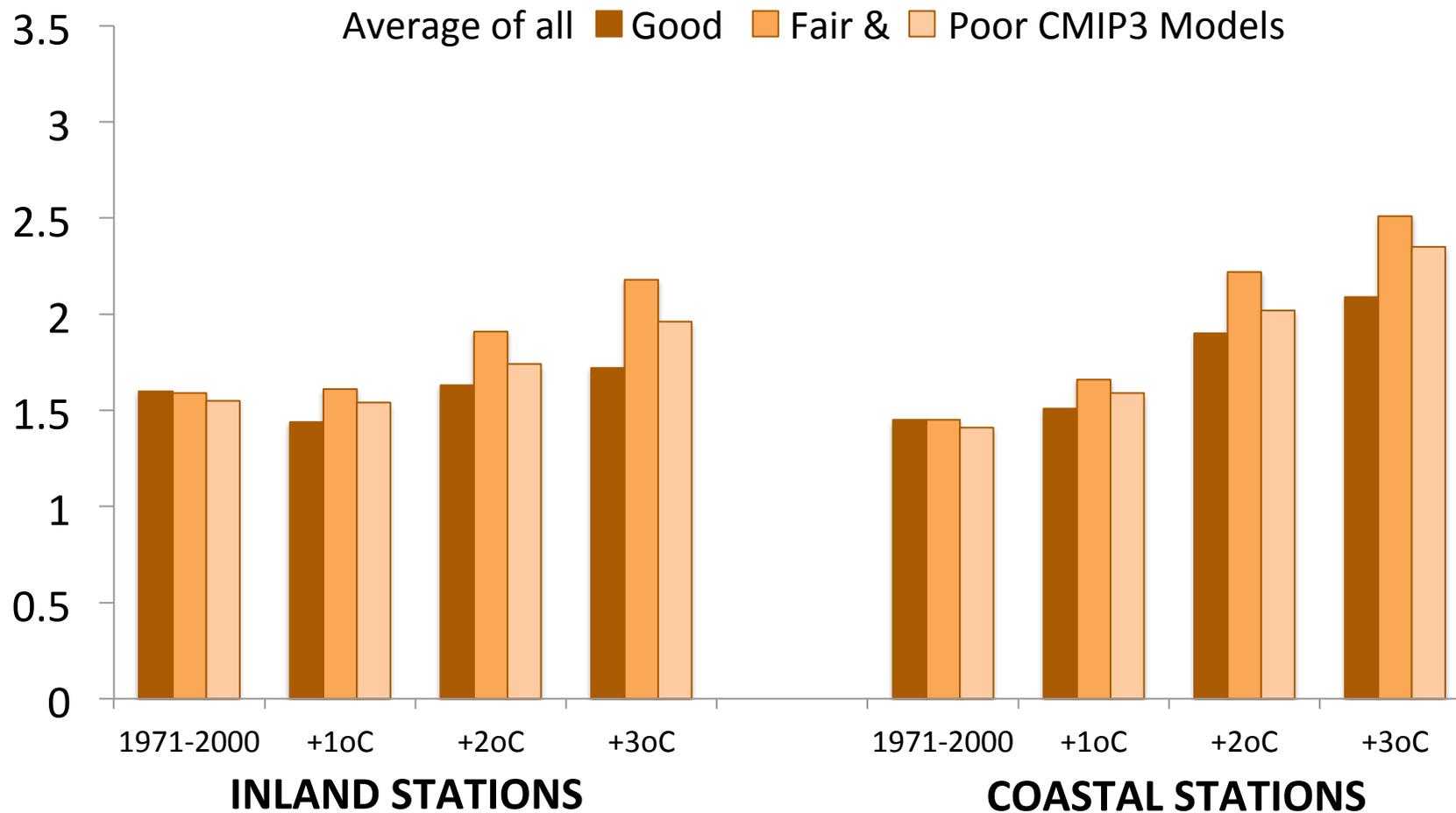
# Standard Deviation of Daily Maximum Temperature Dry Season



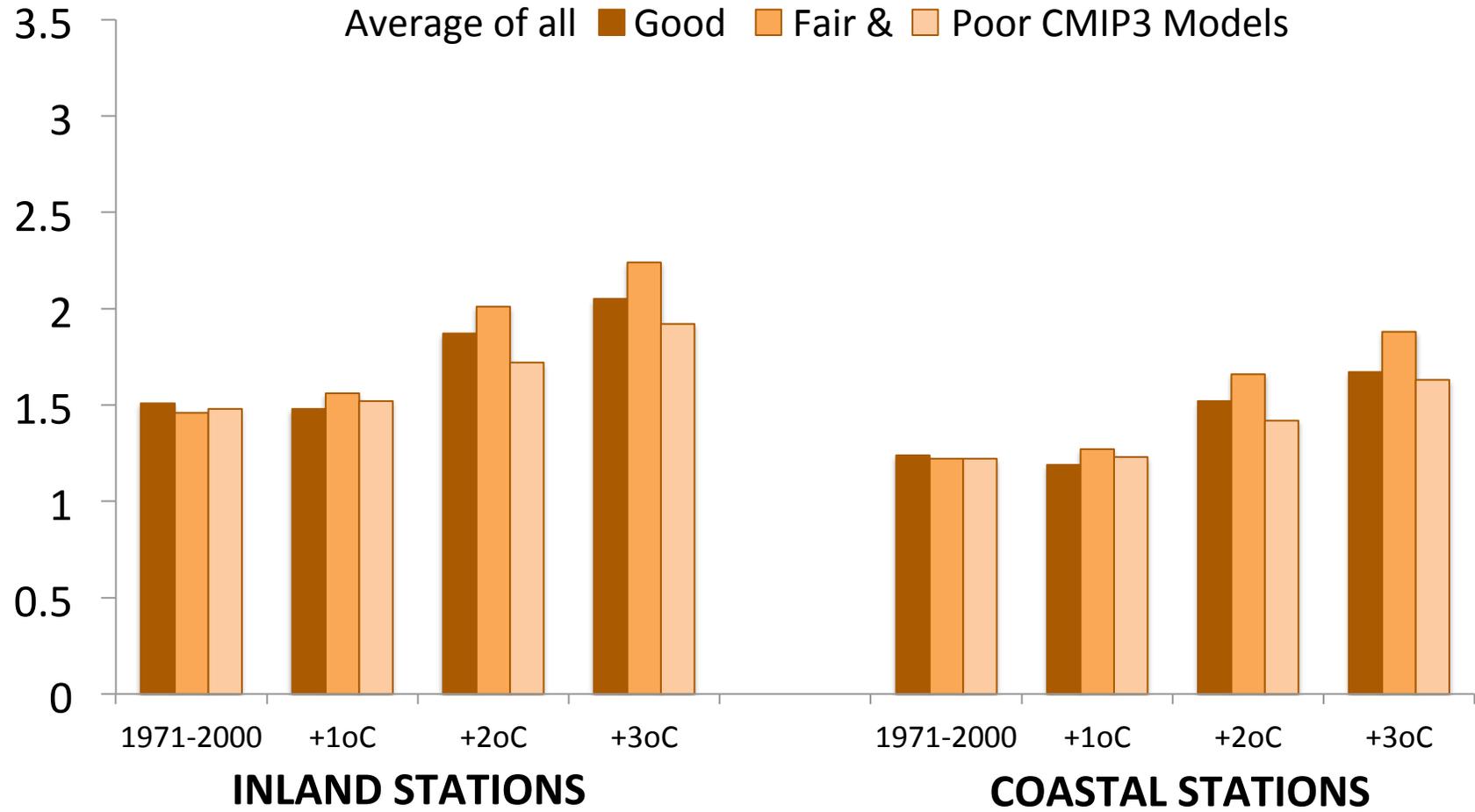
# Standard Deviation of Daily Maximum Temperature Wet Season



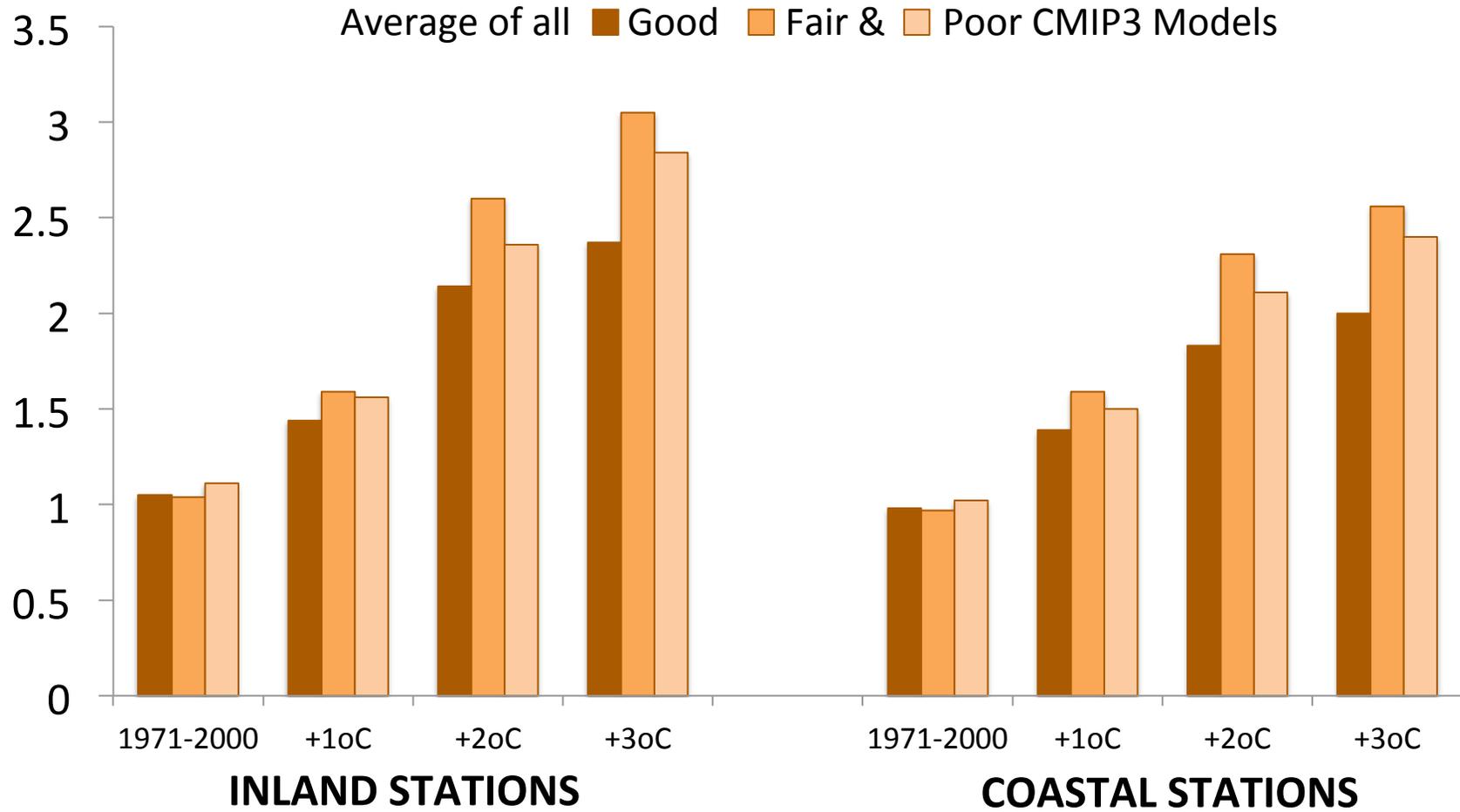
# Standard Deviation of Daily Maximum Temperature Annual



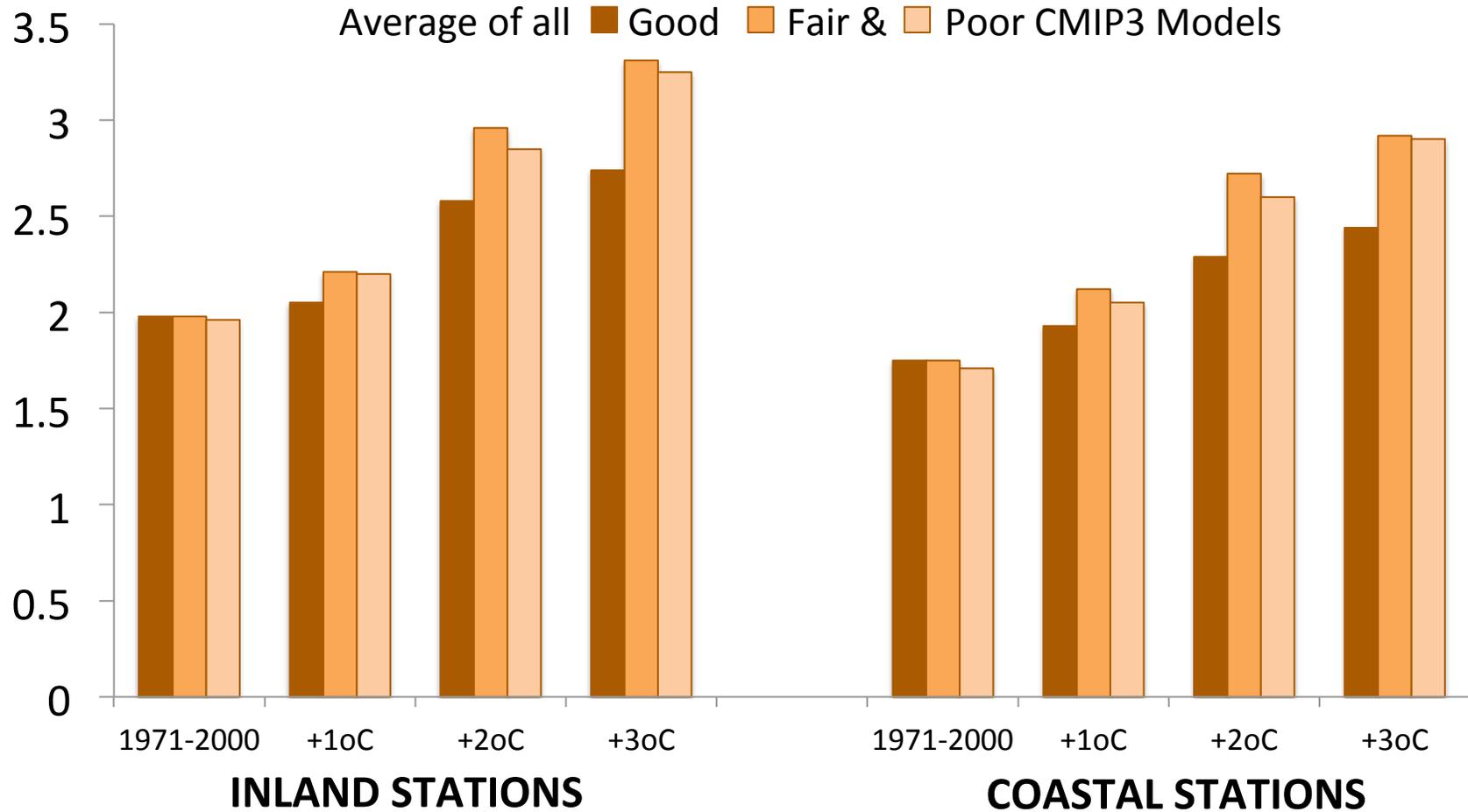
# Standard Deviation of Nighttime Minimum Temperature Dry Season



# Standard Deviation of Nighttime Minimum Temperature Wet Season



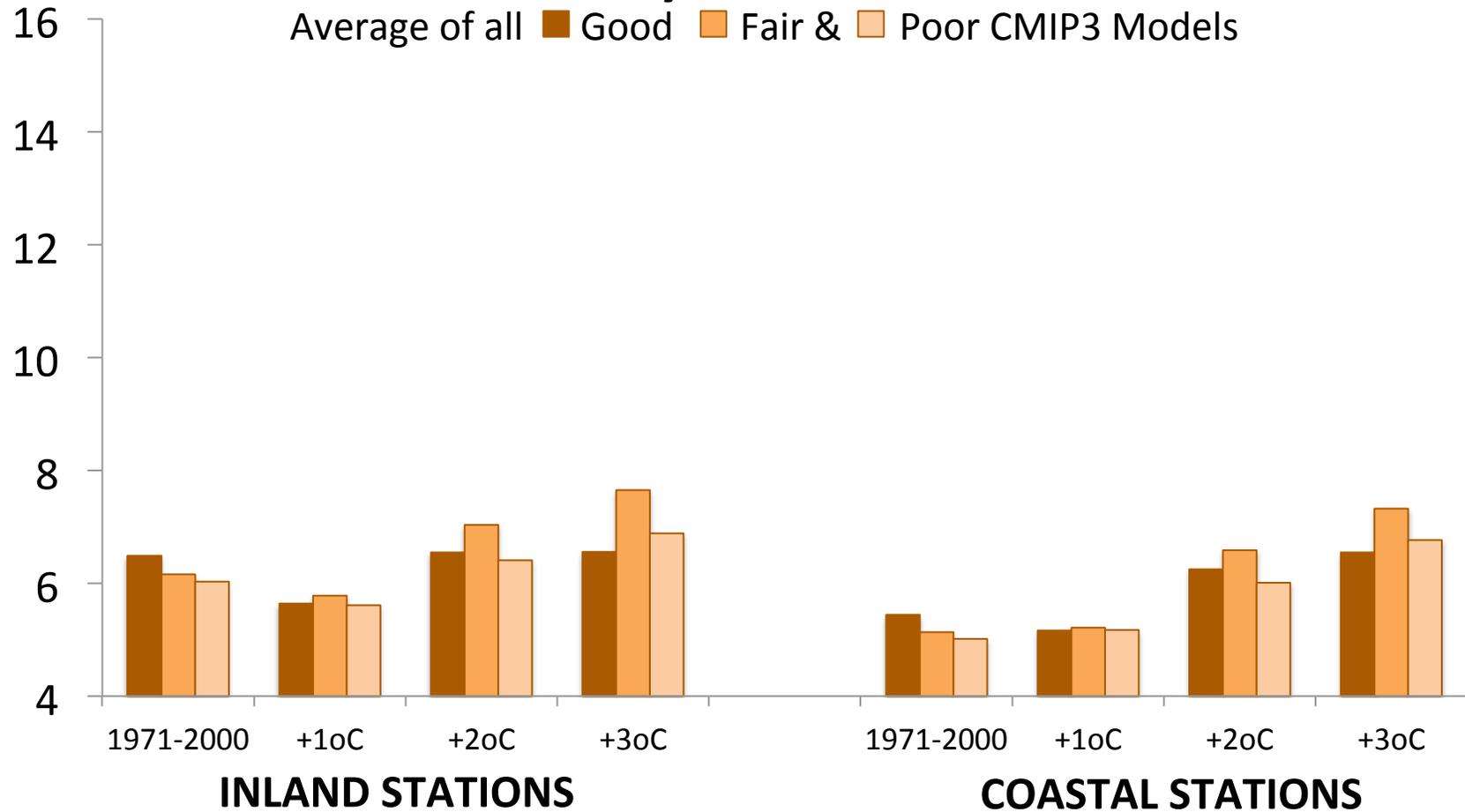
# Standard Deviation of Nighttime Minimum Temperature Annual



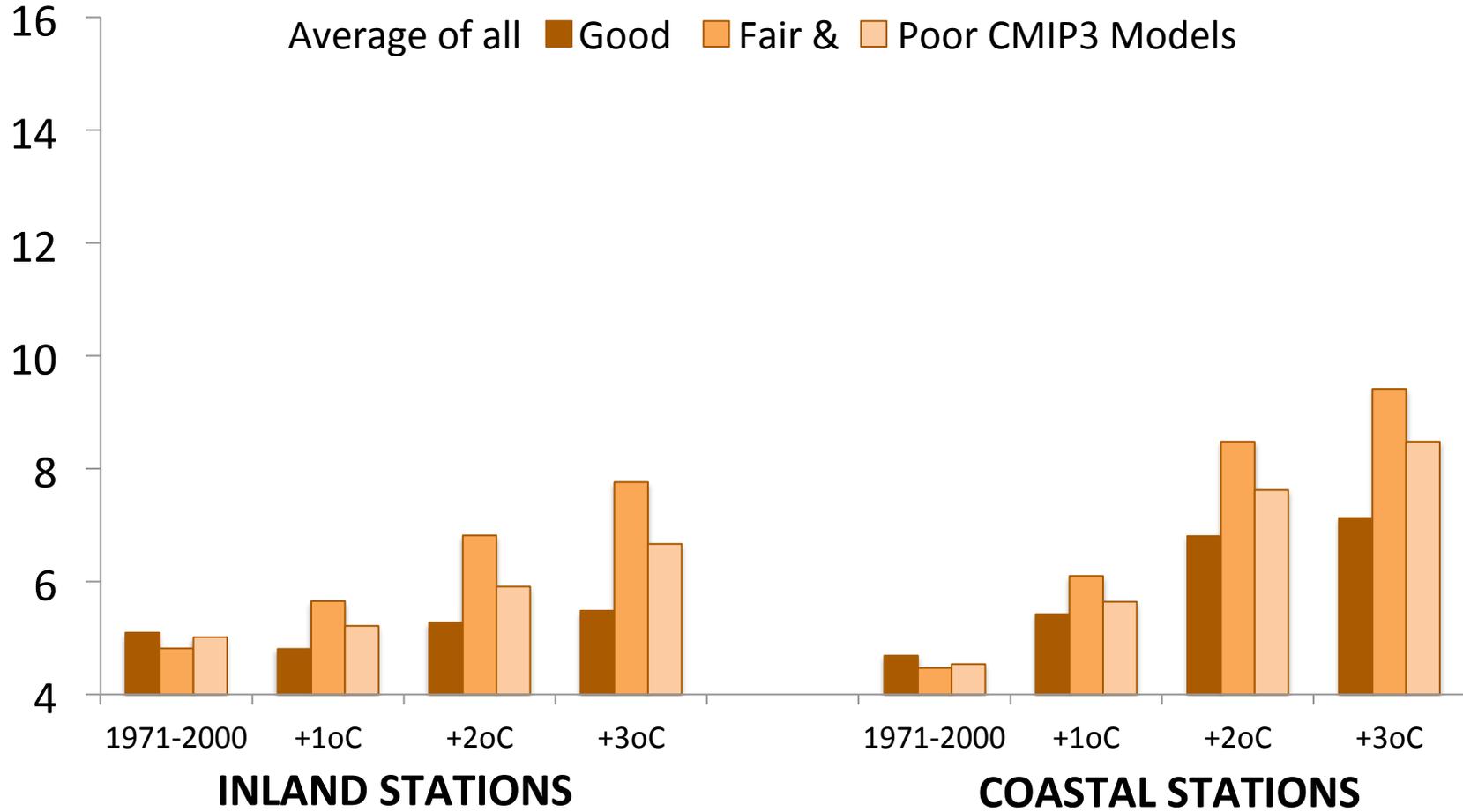
# Range in Daily Maximum Temperature

## Dry Season

Average of all **Good** **Fair &** **Poor** CMIP3 Models



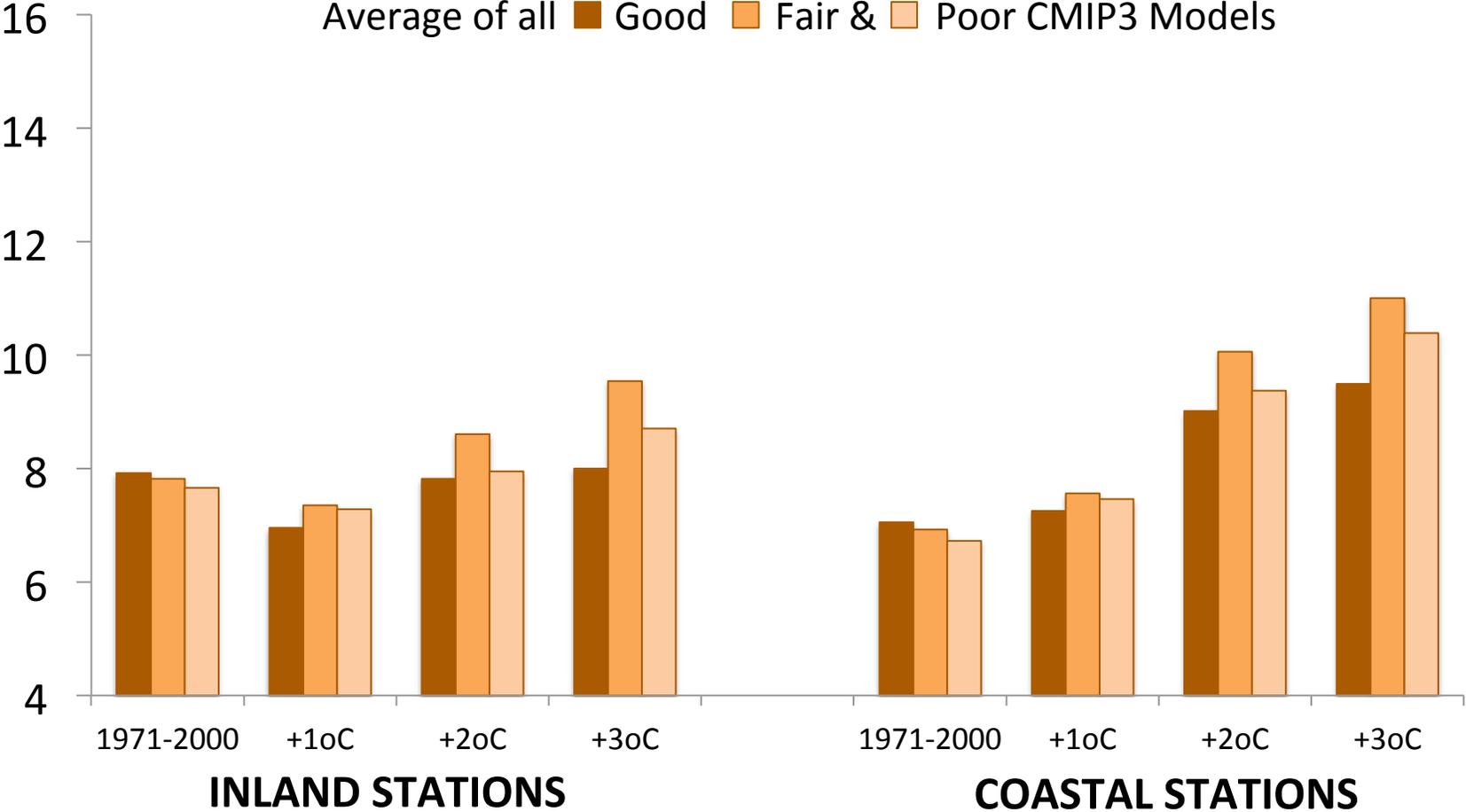
# Range in Daily Maximum Temperature Wet Season



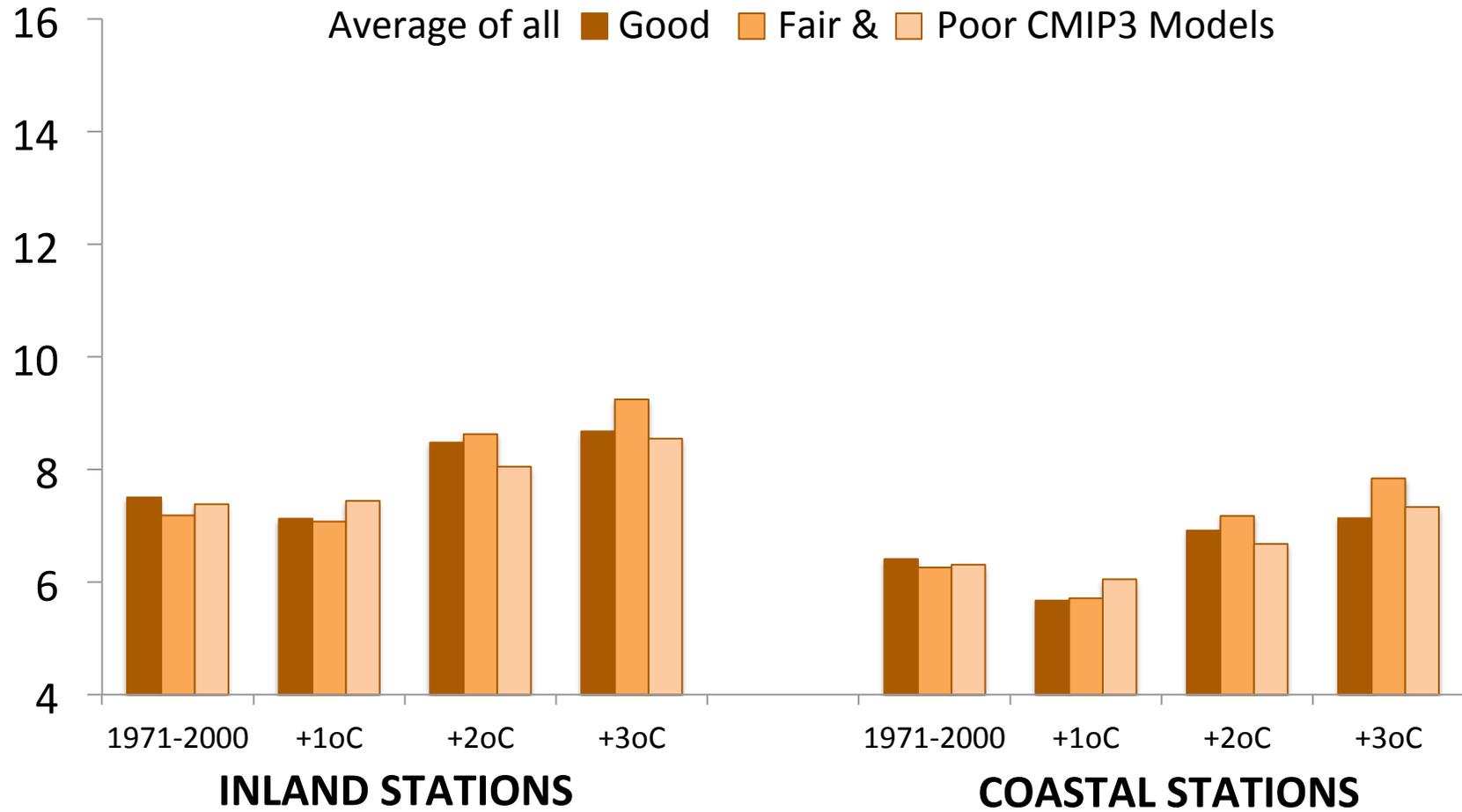
# Range in Daily Maximum Temperature

## Annual

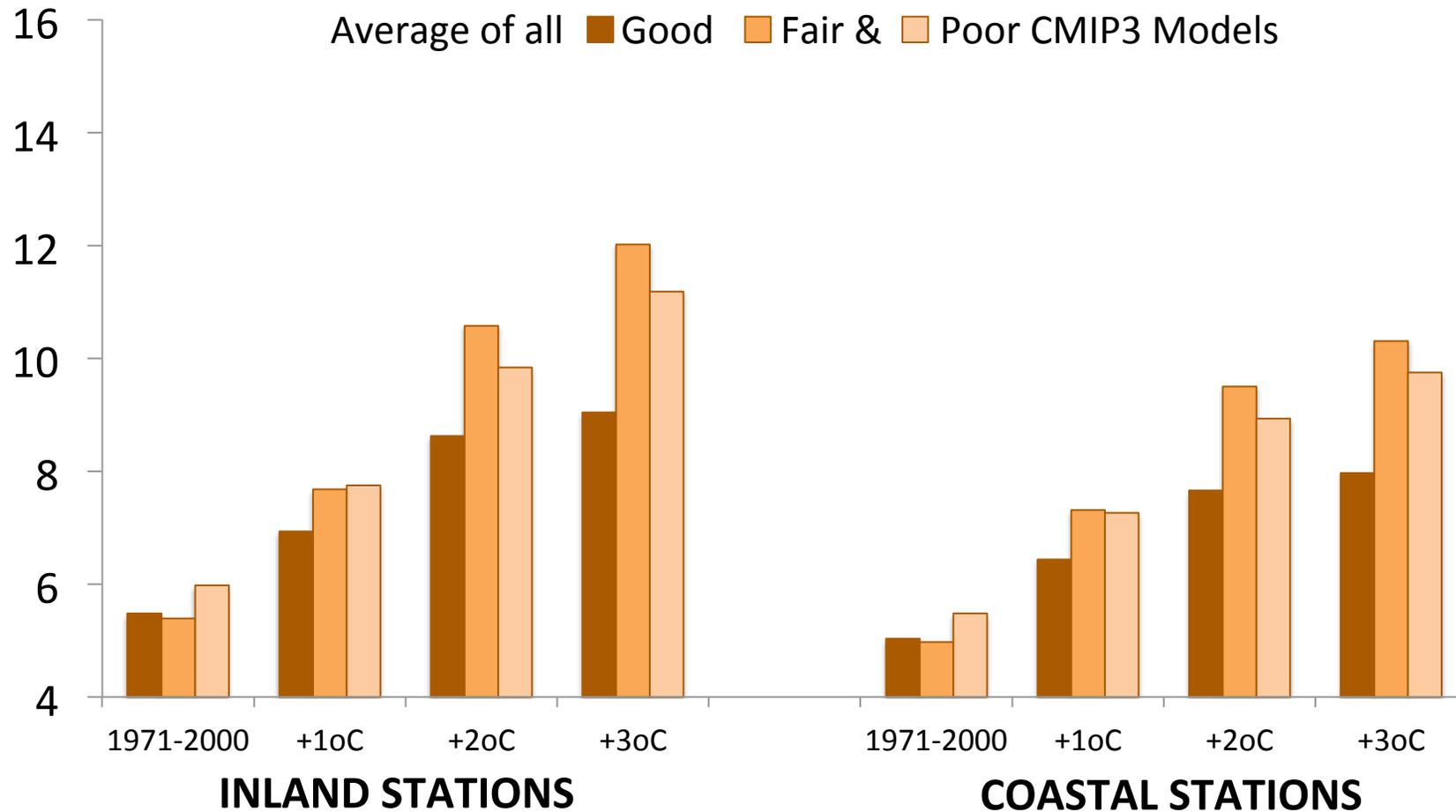
Average of all **Good** **Fair &** **Poor** CMIP3 Models



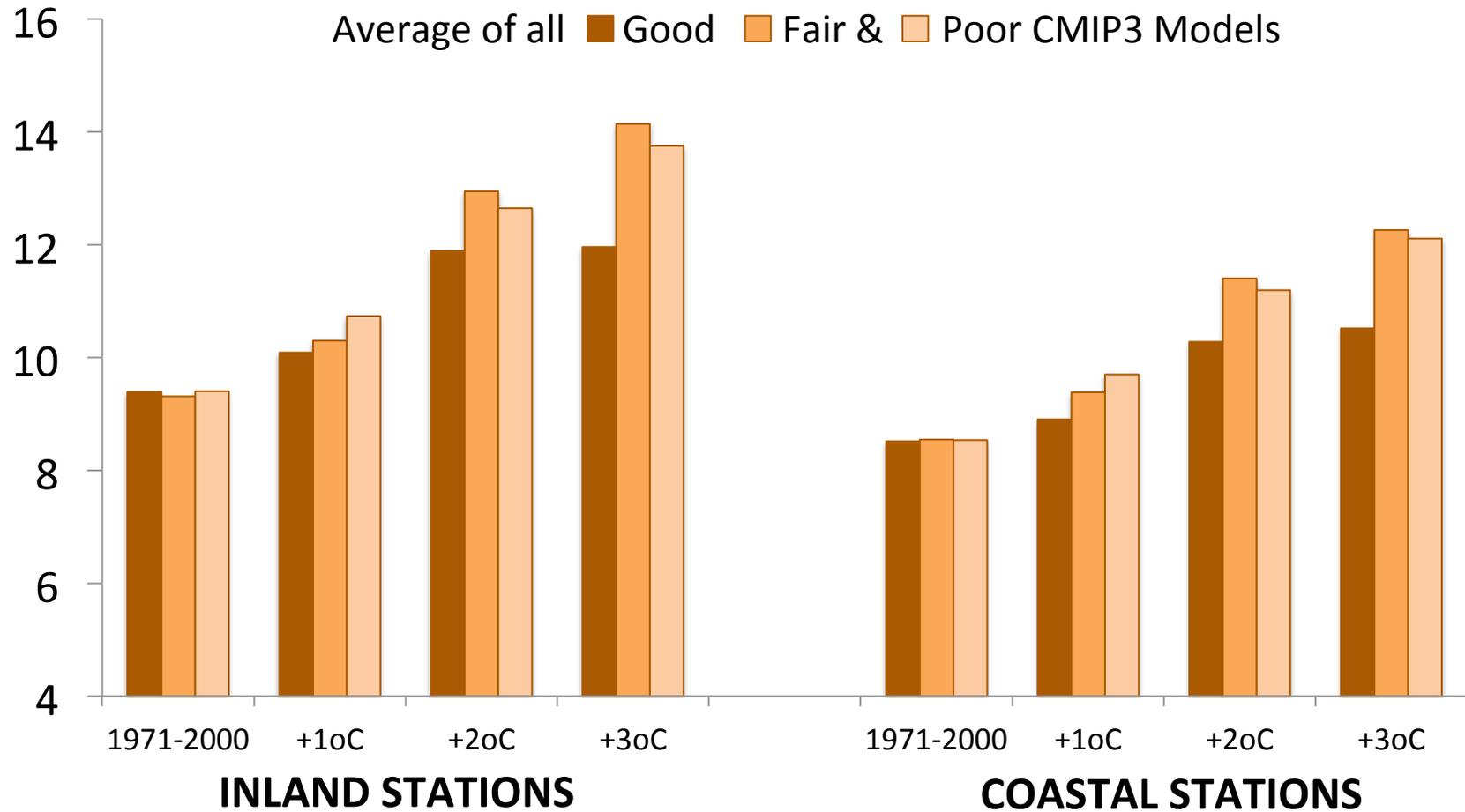
# Range in Nighttime Minimum Temperature Dry Season



# Range in Nighttime Minimum Temperature Wet Season

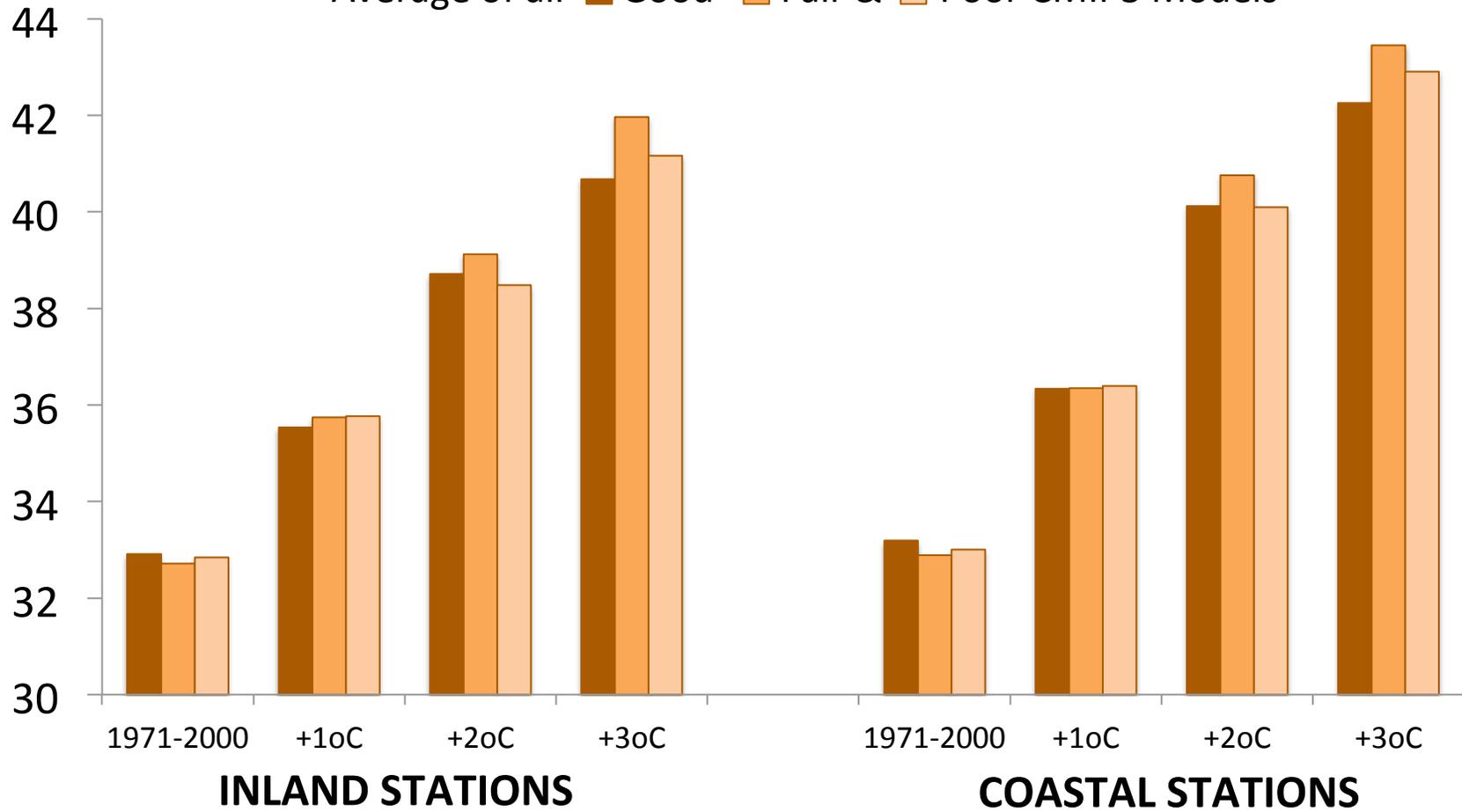


# Range in Nighttime Minimum Temperature Annual

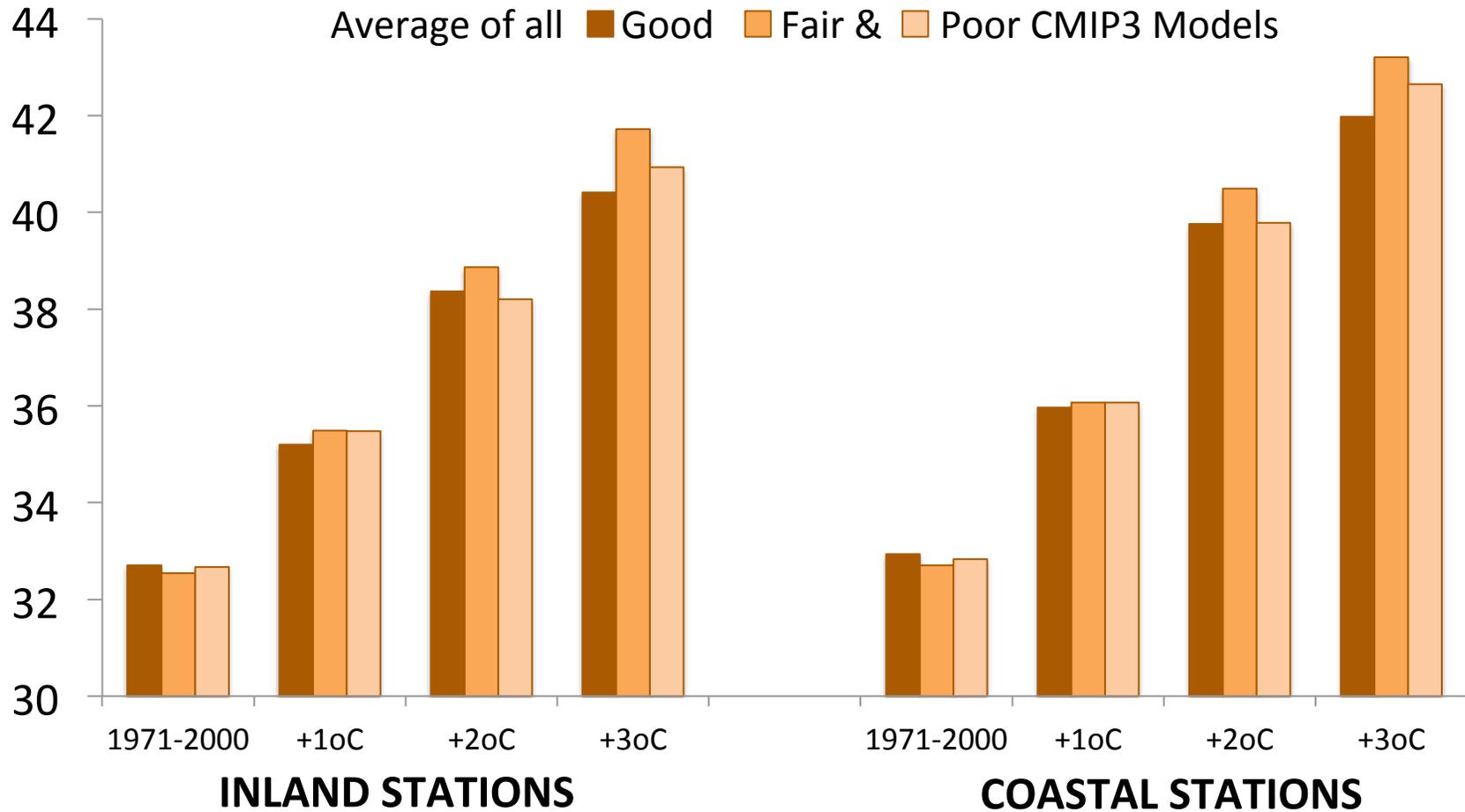


# Daily Maximum Temperature on Hottest Day of Year

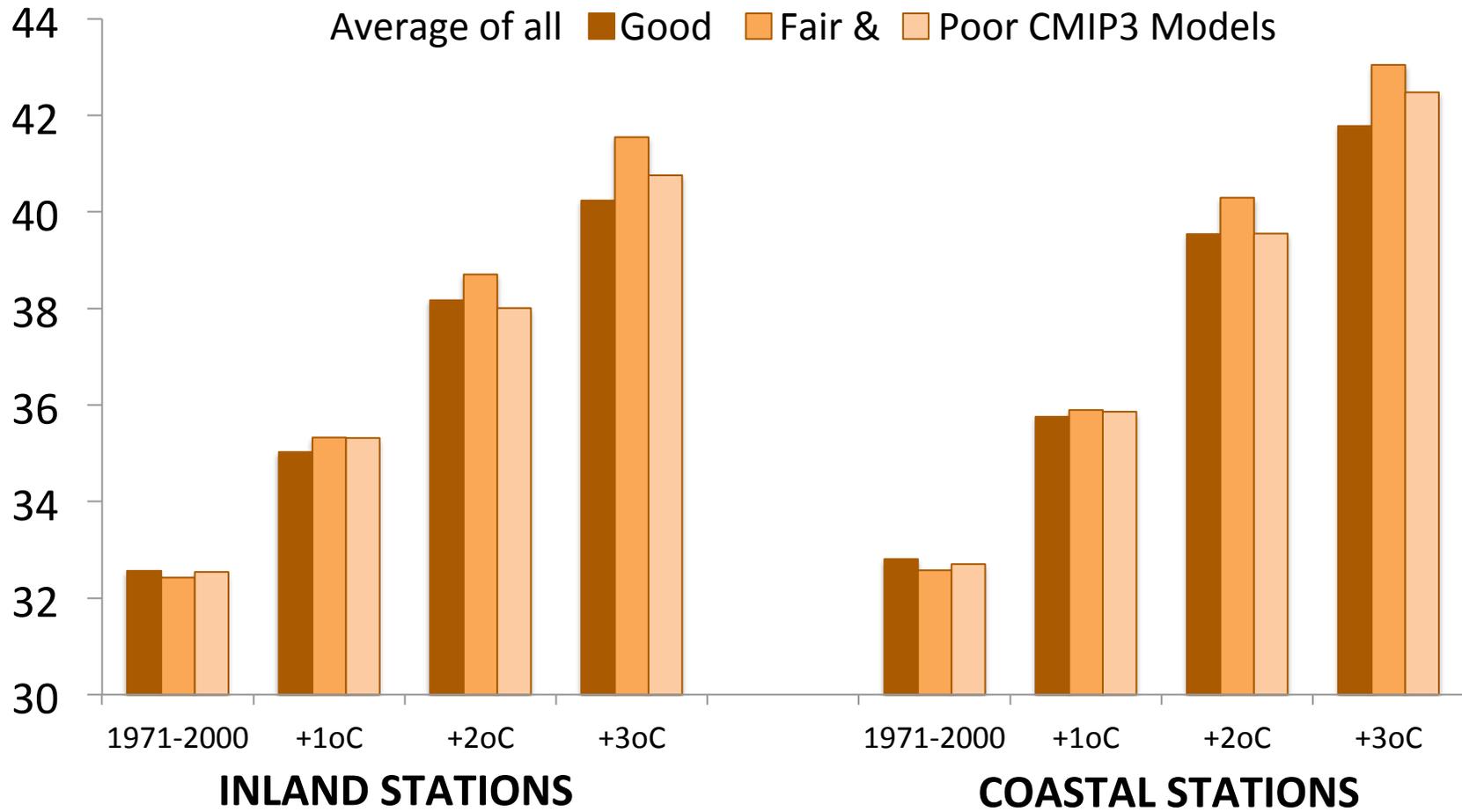
Average of all **Good** **Fair &** **Poor CMIP3 Models**



# Daily Maximum Temperature on the Hottest 3-Day Period of the Year

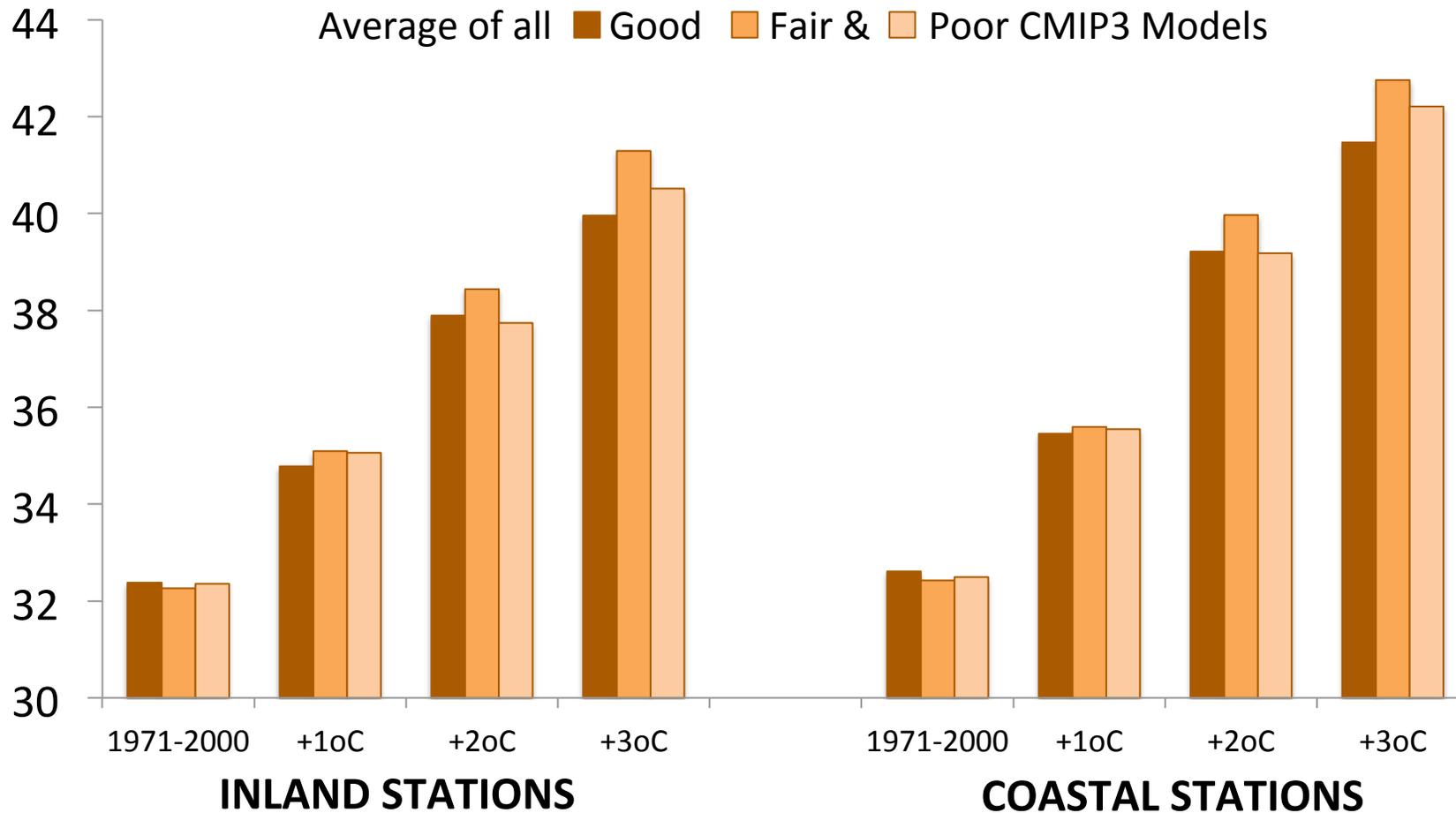


# Daily Maximum Temperature on the Hottest 5-Day Period of the Year

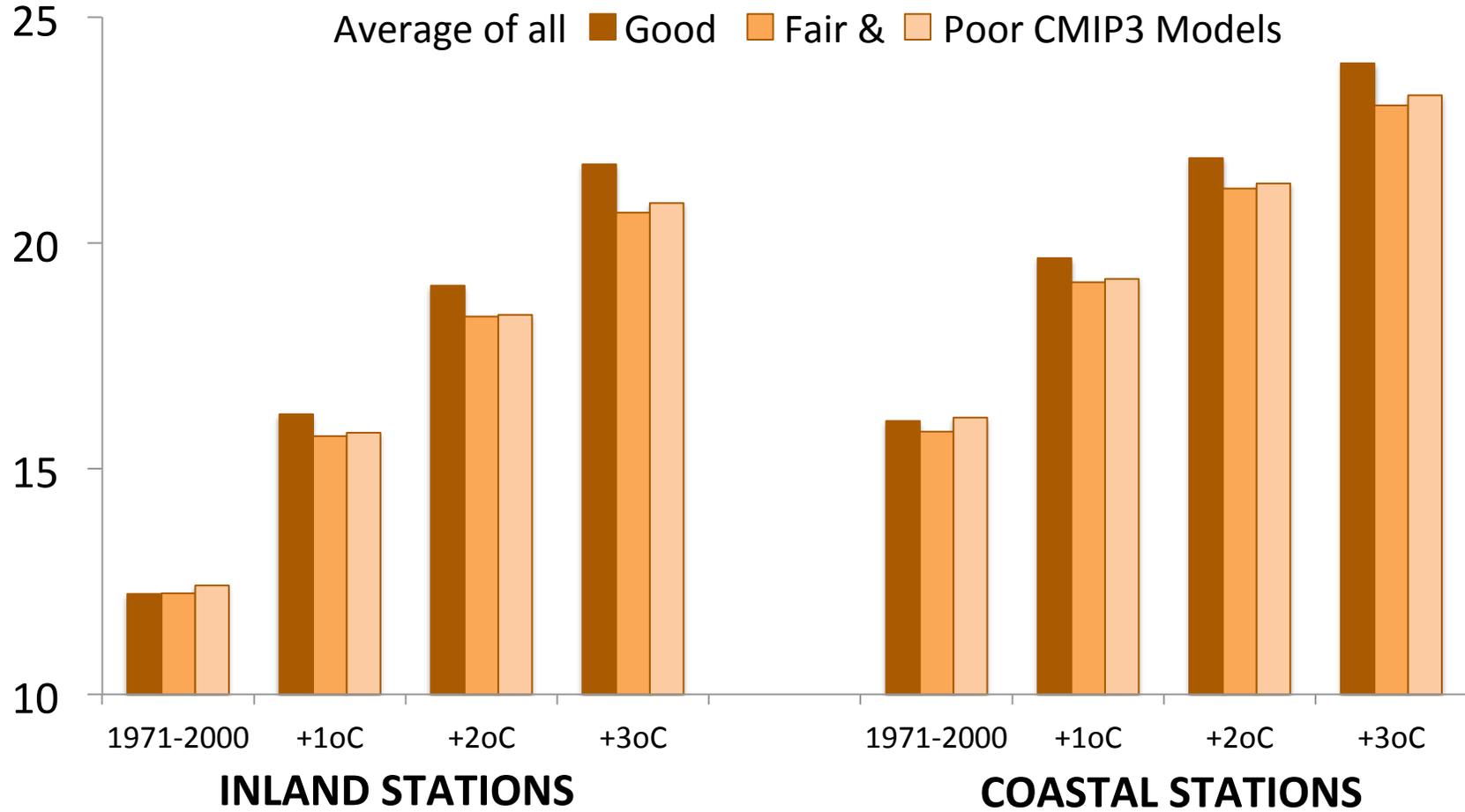


# Daily Maximum Temperature on the Hottest 10-Day Period of the Year

Average of all **Good** **Fair &** **Poor** CMIP3 Models

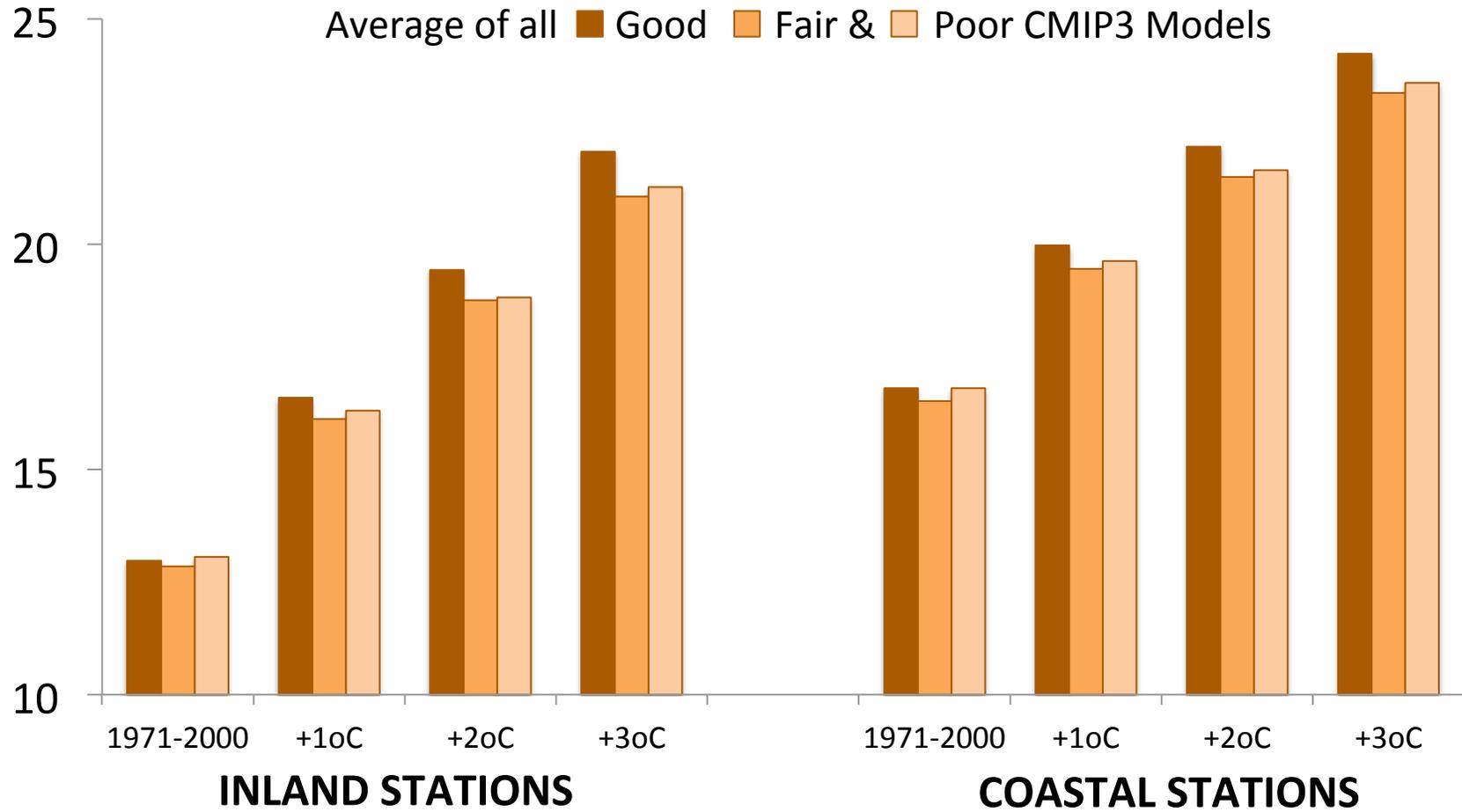


# Nighttime Minimum Temperature on the Coldest Day of the Year

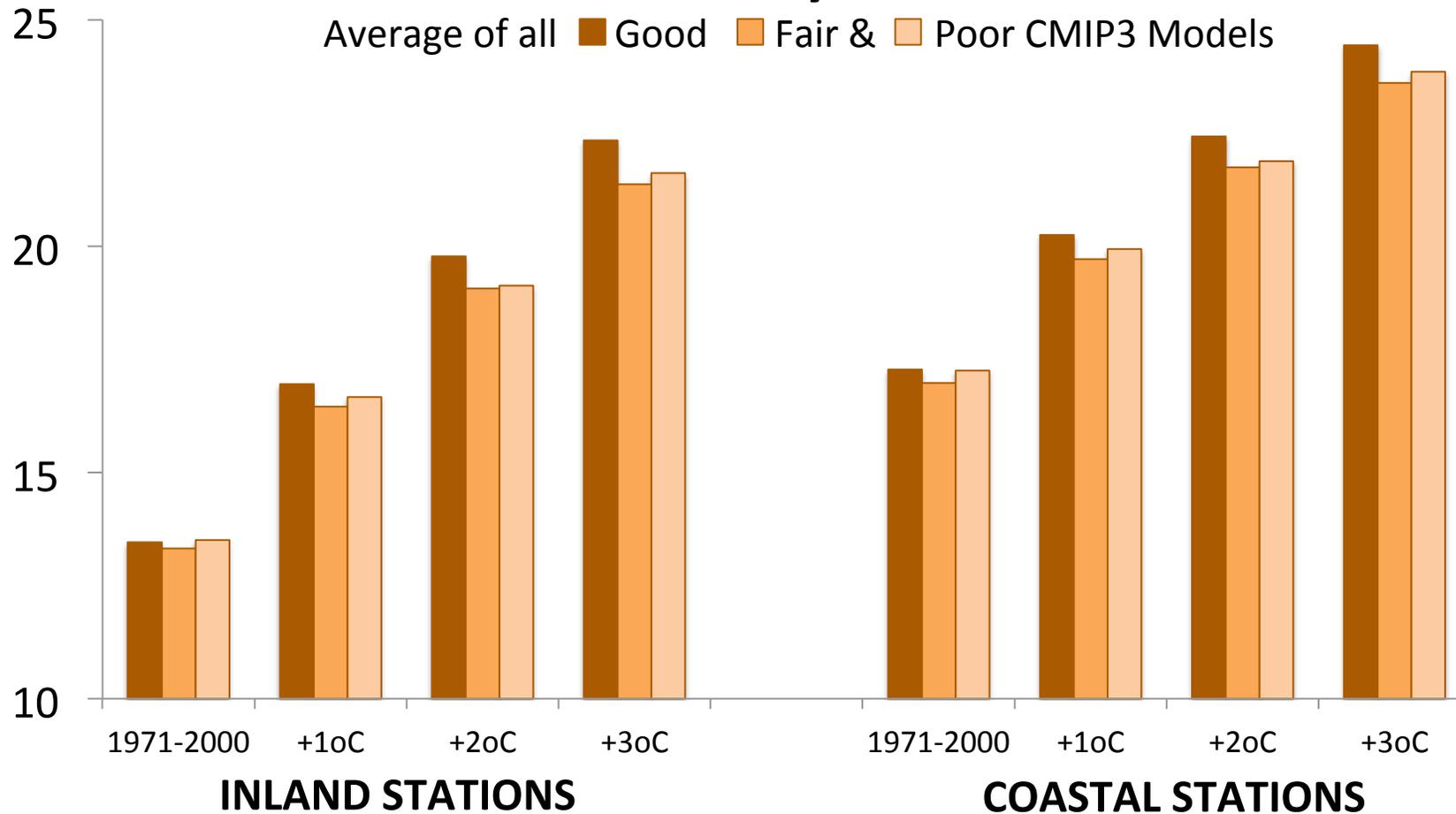


# Nighttime Minimum Temperature on the Coldest 3-Day Period of the Year

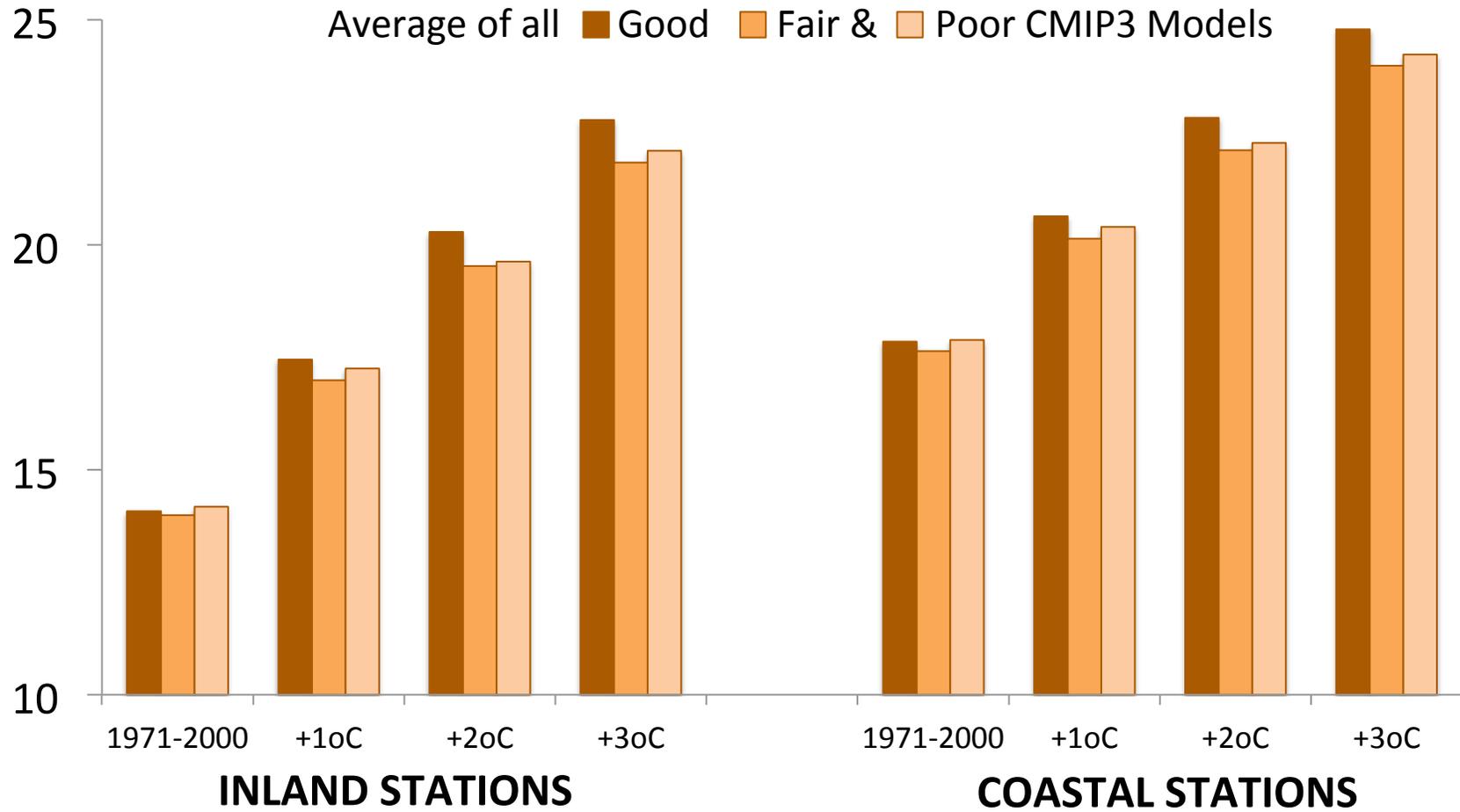
Average of all **Good** **Fair &** **Poor** CMIP3 Models



# Nighttime Minimum Temperature on the Coldest 5-Day Period of the Year

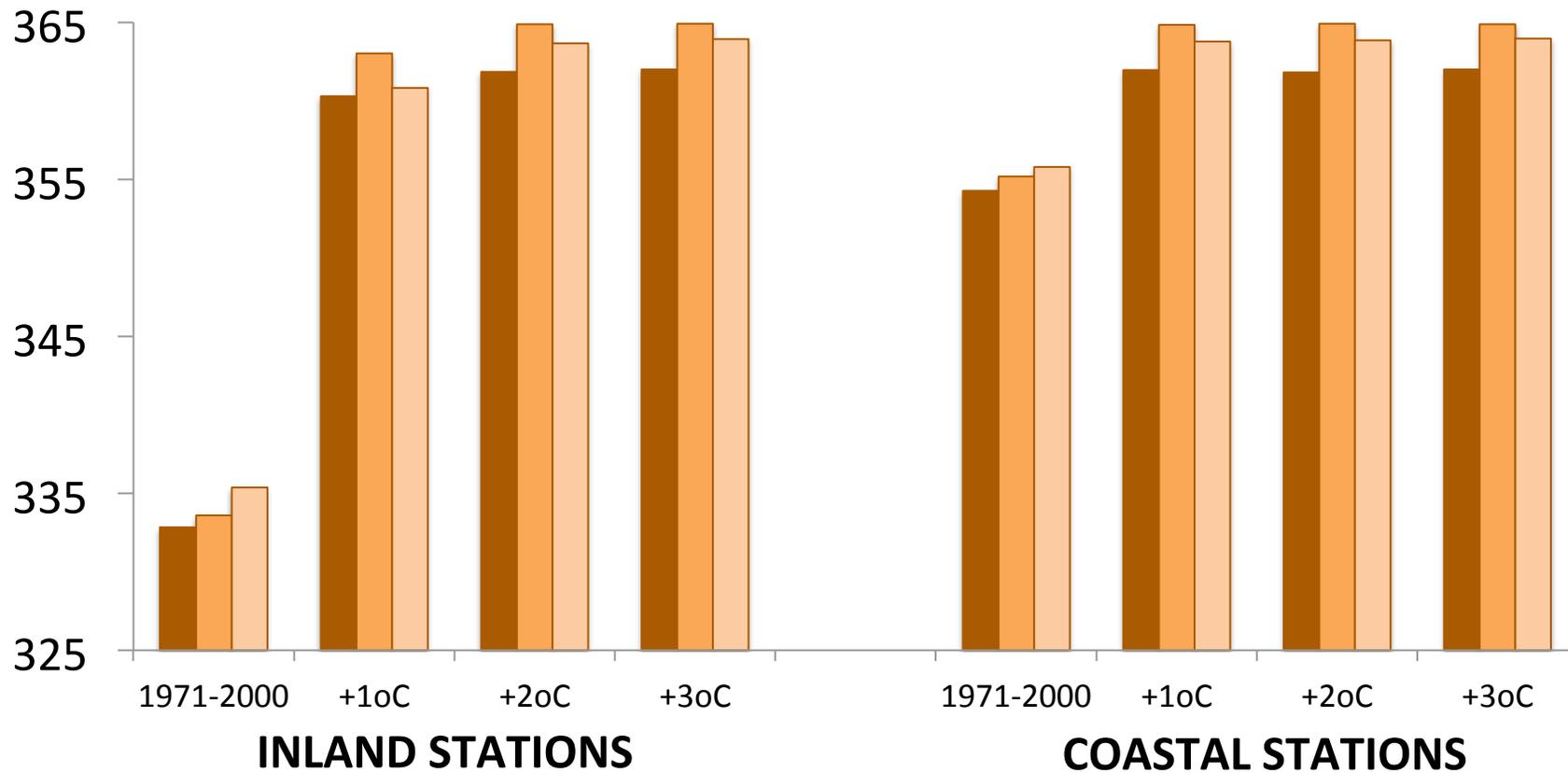


# Nighttime Minimum Temperature on the Coldest 10-Day Period of the Year



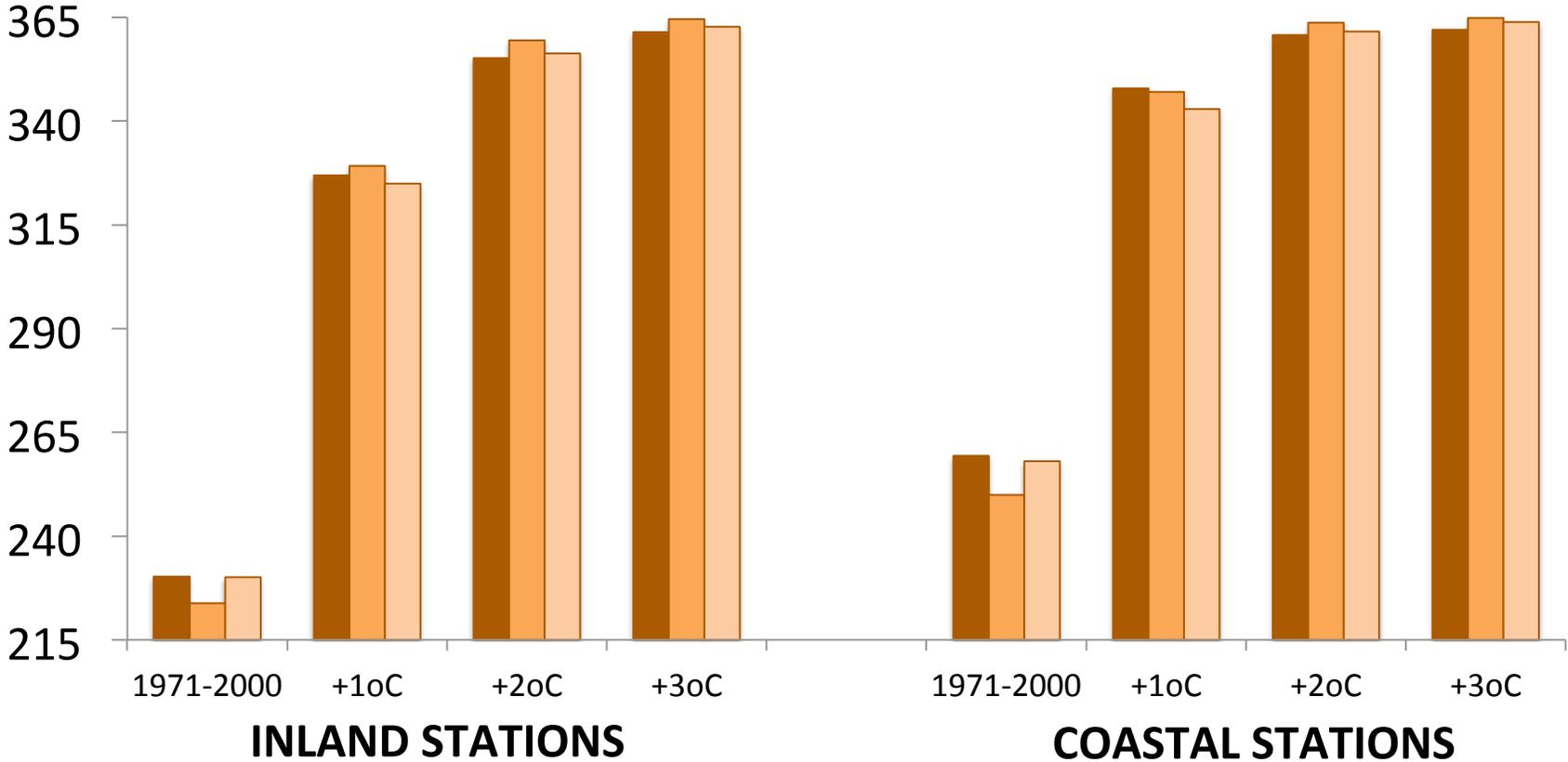
# Days per Year with Daily Maximum Temperature > 80oF

Average of all **Good** **Fair &** **Poor** CMIP3 Models

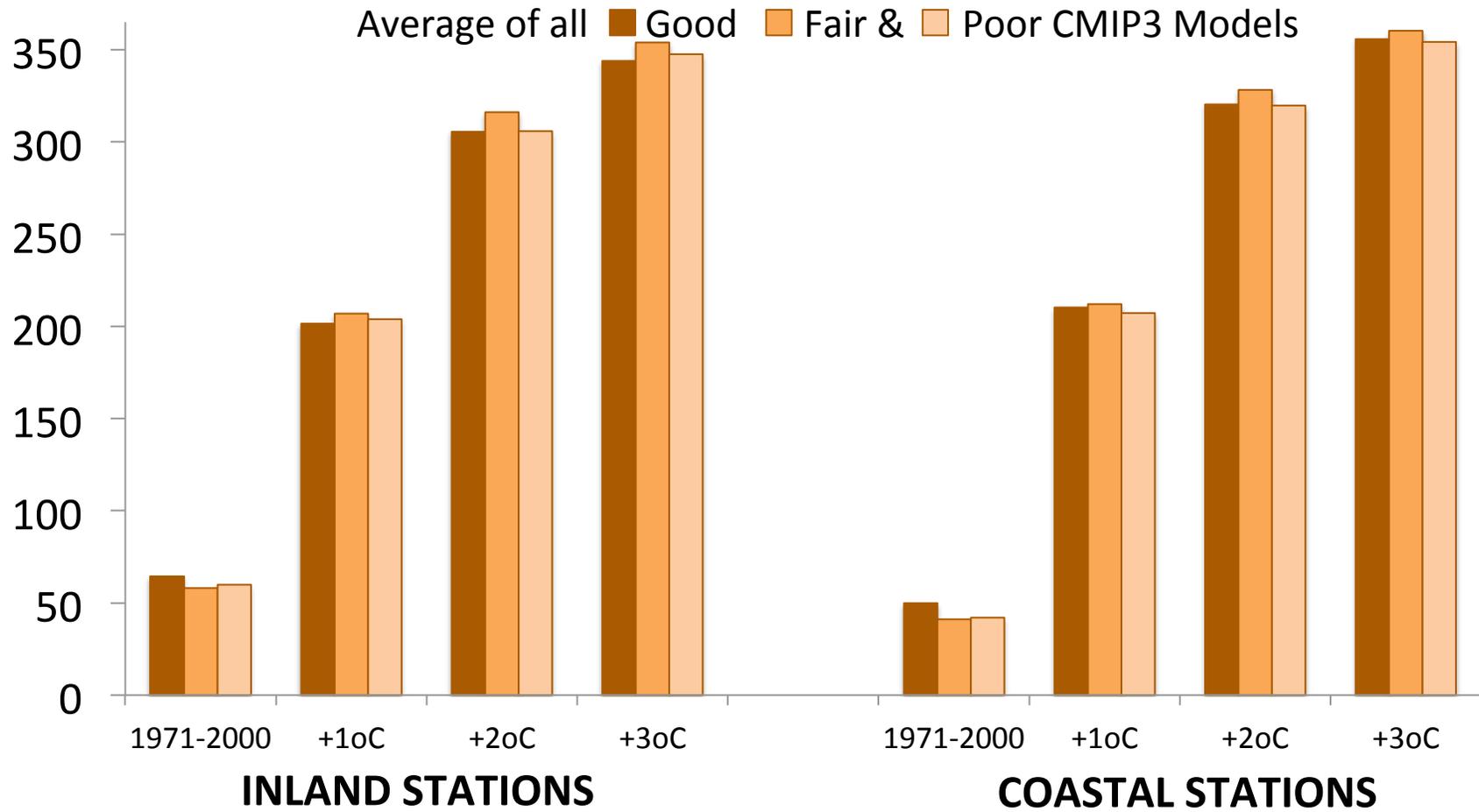


# Days per Year with Daily Maximum Temperature > 85oF

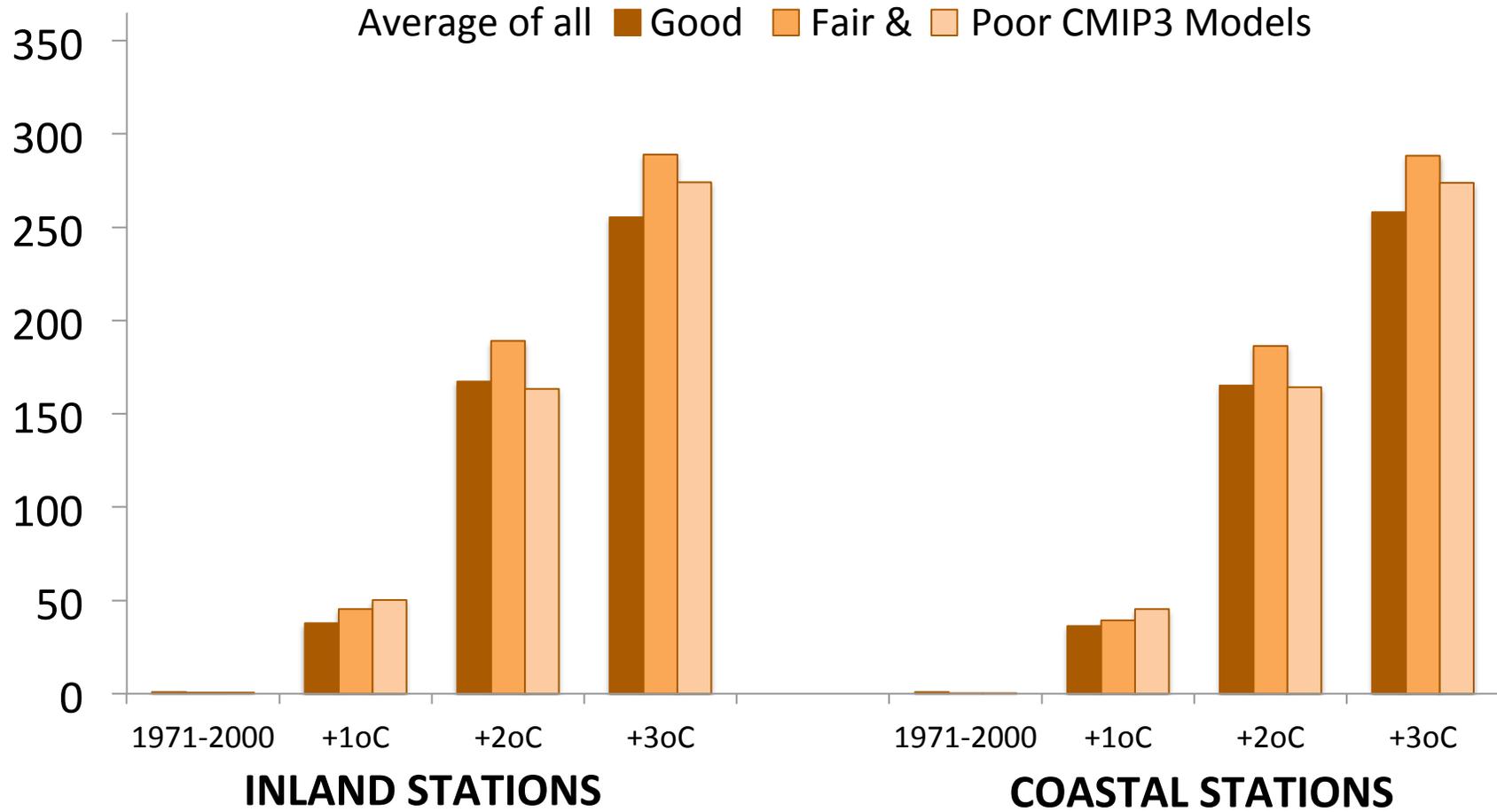
Average of all **Good** **Fair &** **Poor** CMIP3 Models



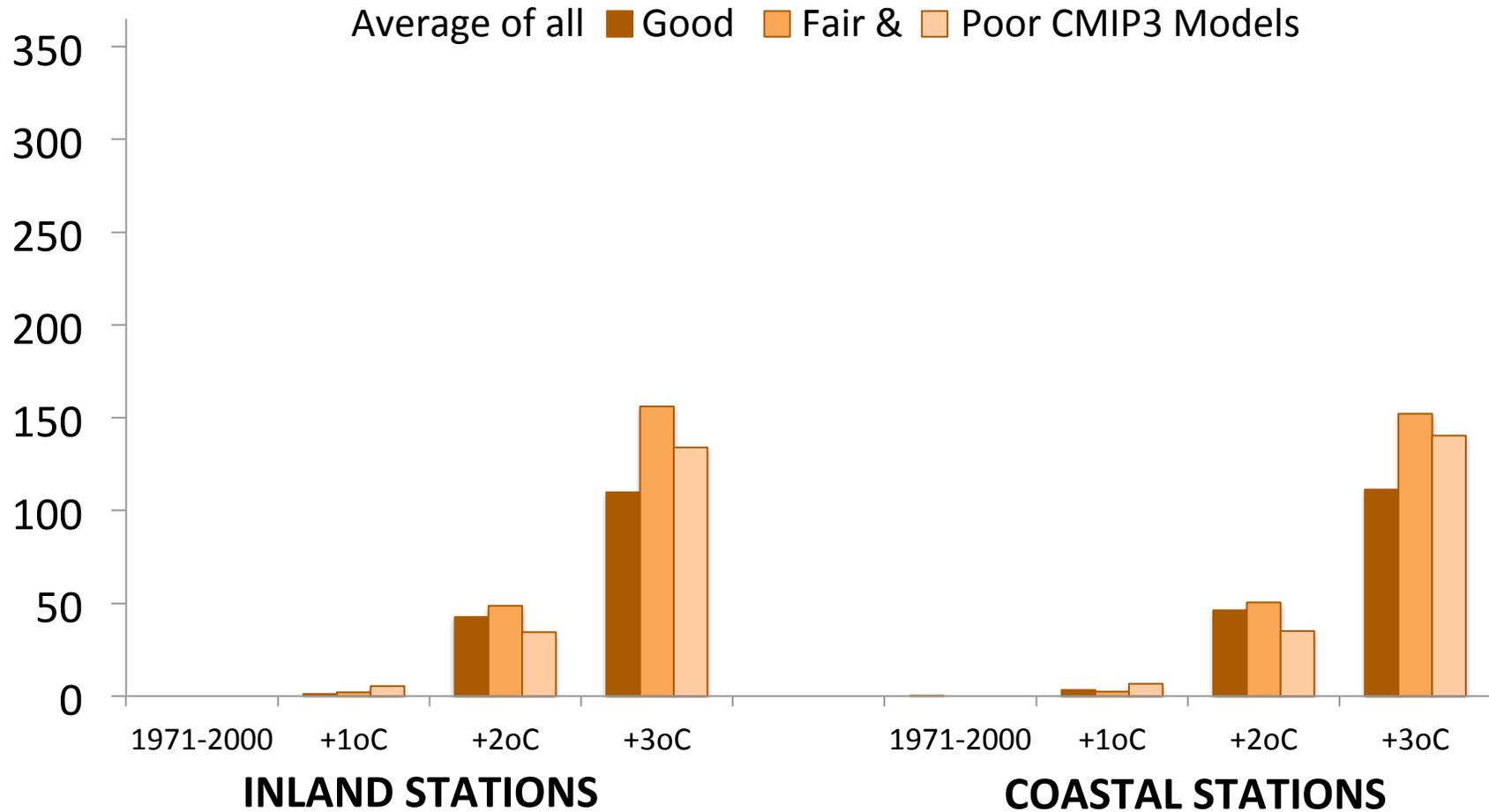
# Days per Year with Daily Maximum Temperature > 90oF



# Days per Year with Daily Maximum Temperature > 95oF

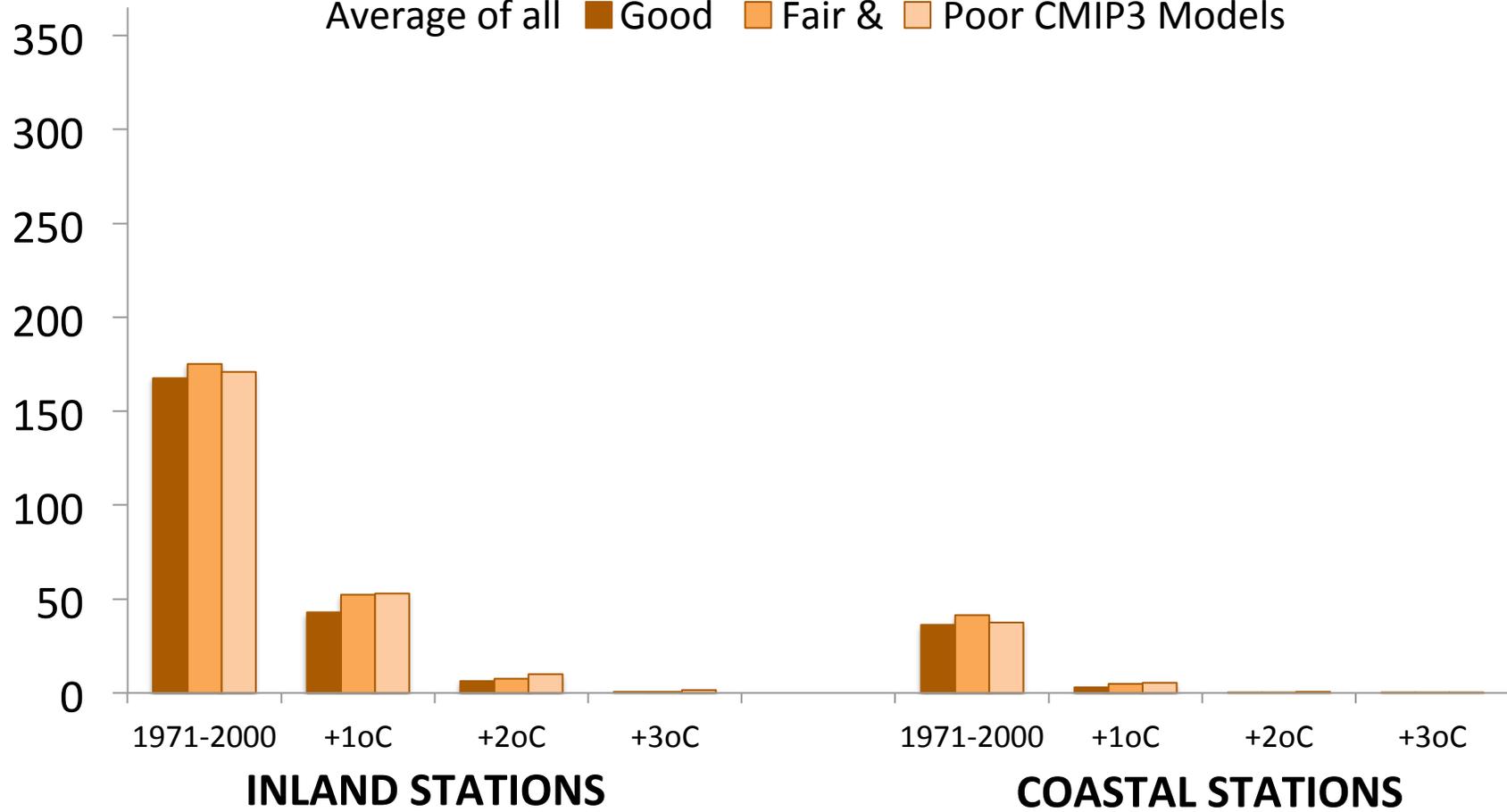


# Days per Year with Daily Maximum Temperature > 100oF

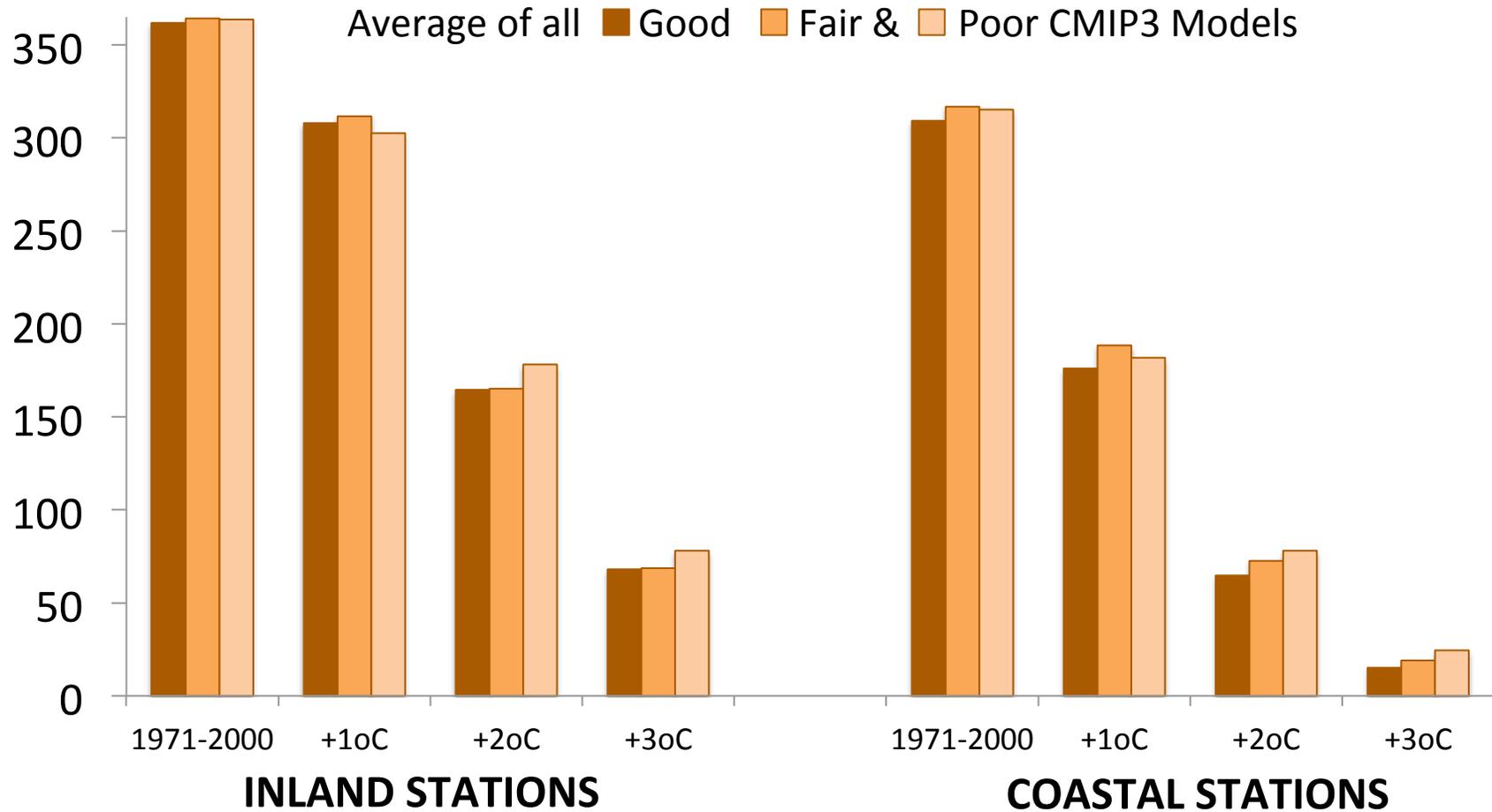


# Days per Year with Nighttime Minimum Temperature < 65oF

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models

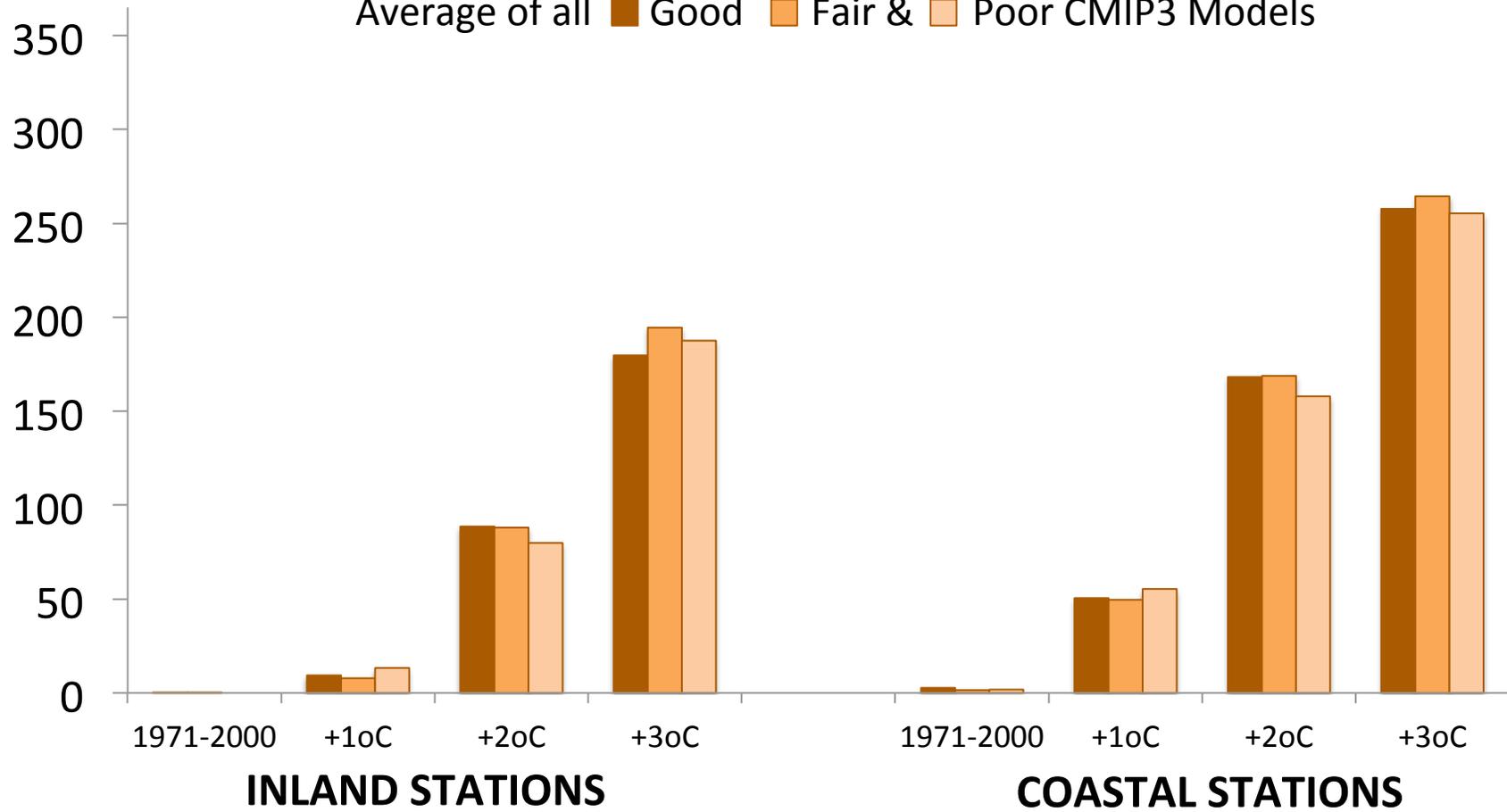


# Days per Year with Nighttime Minimum Temperature < 75oF



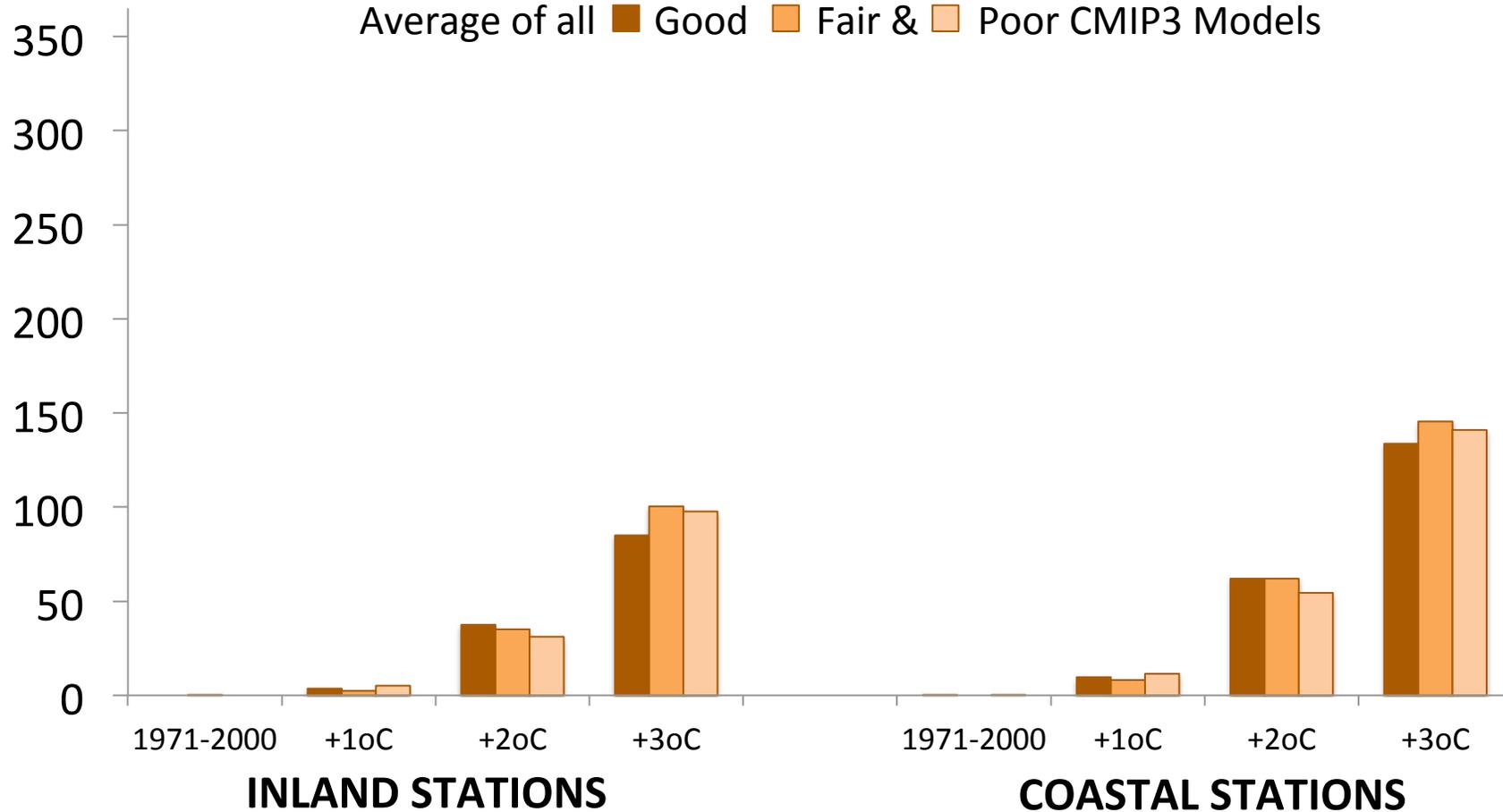
# Days per Year with Nighttime Minimum Temperature > 80oF

Average of all **Good** **Fair &** **Poor** CMIP3 Models



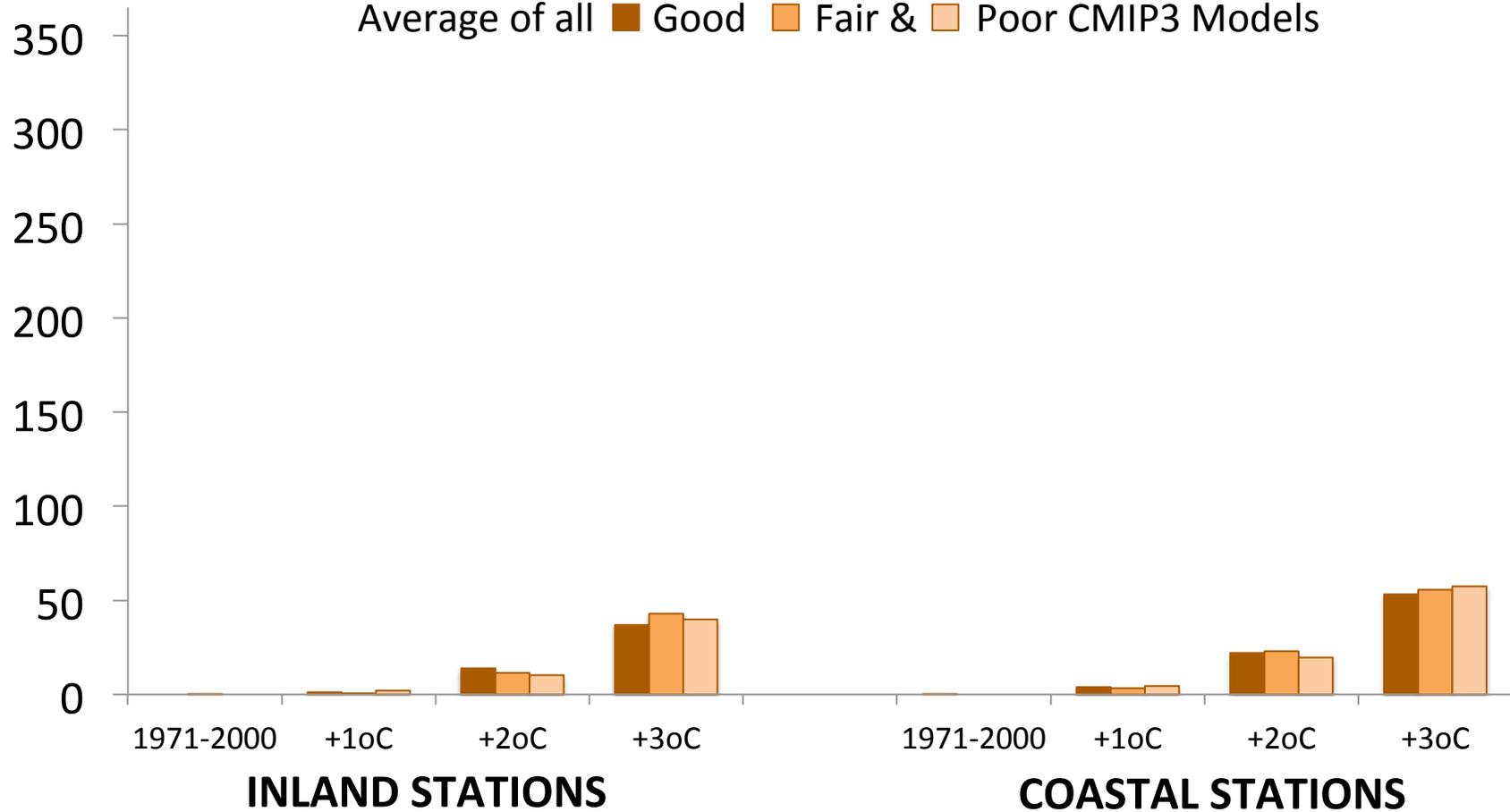
# Days per Year with Nighttime Minimum Temperature > 85oF

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models

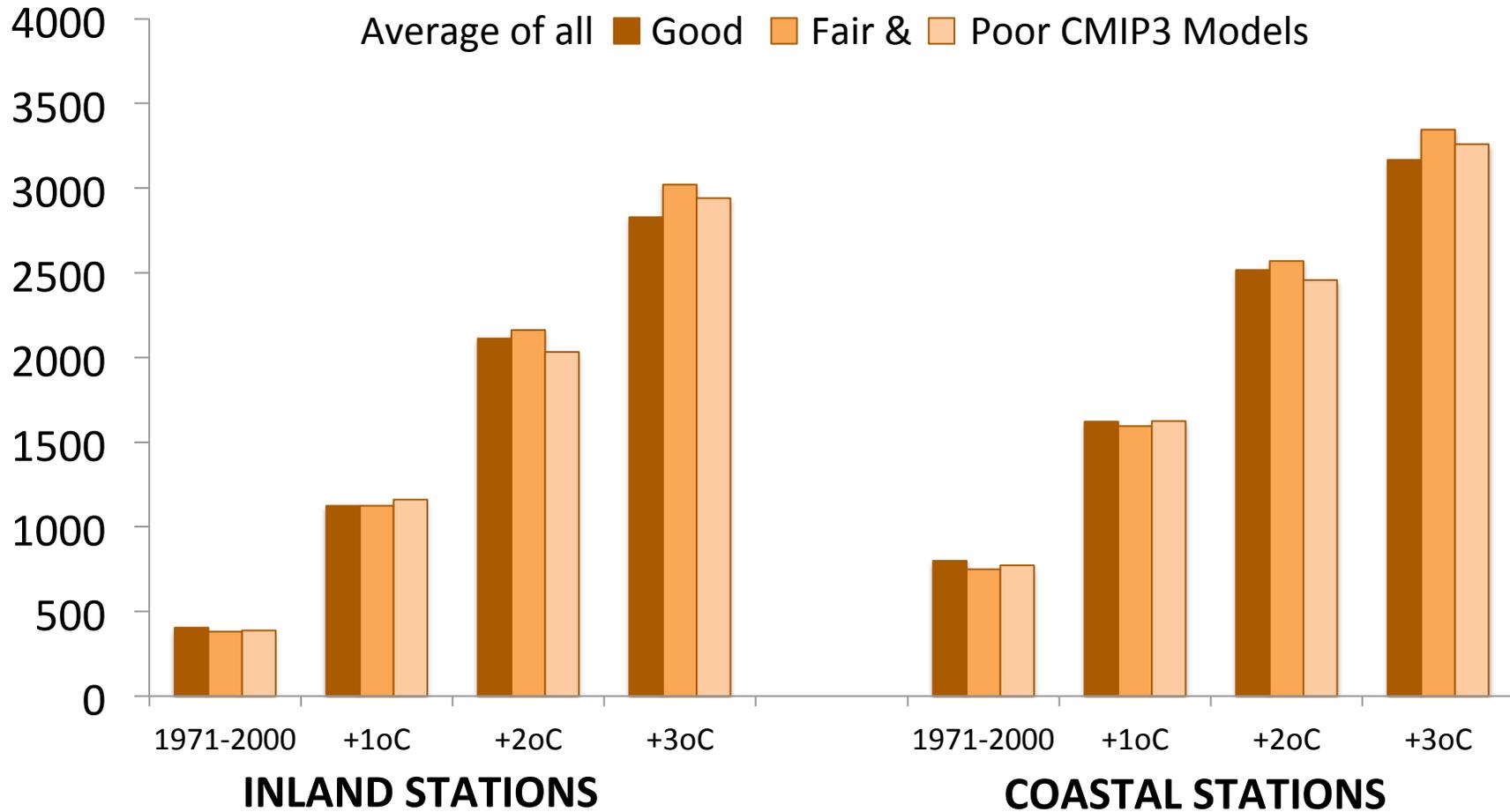


# Days per Year with Nighttime Minimum Temperature > 90oF

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models

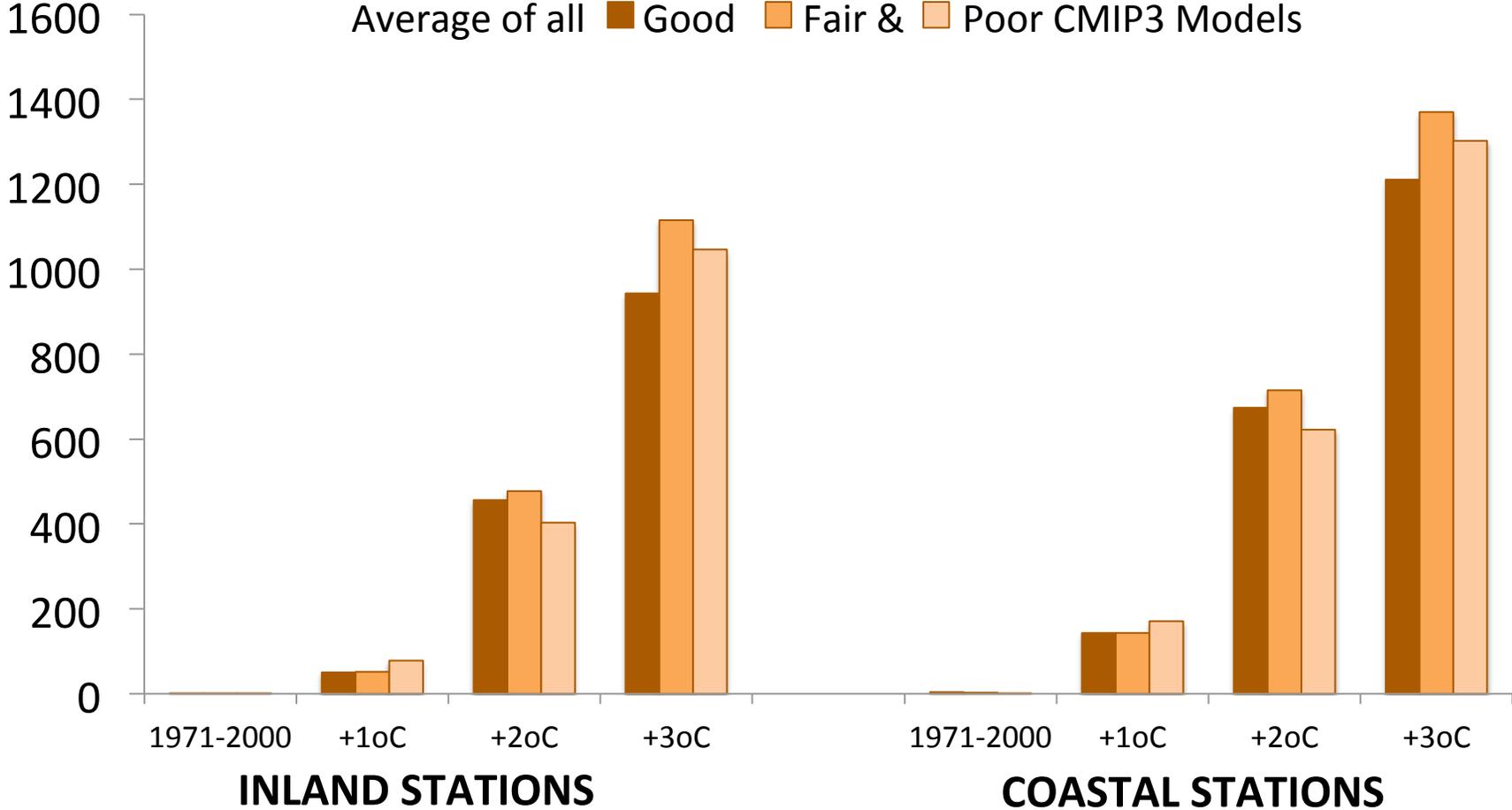


# Annual Cumulative Cooling Degree-Days 75oF Threshold



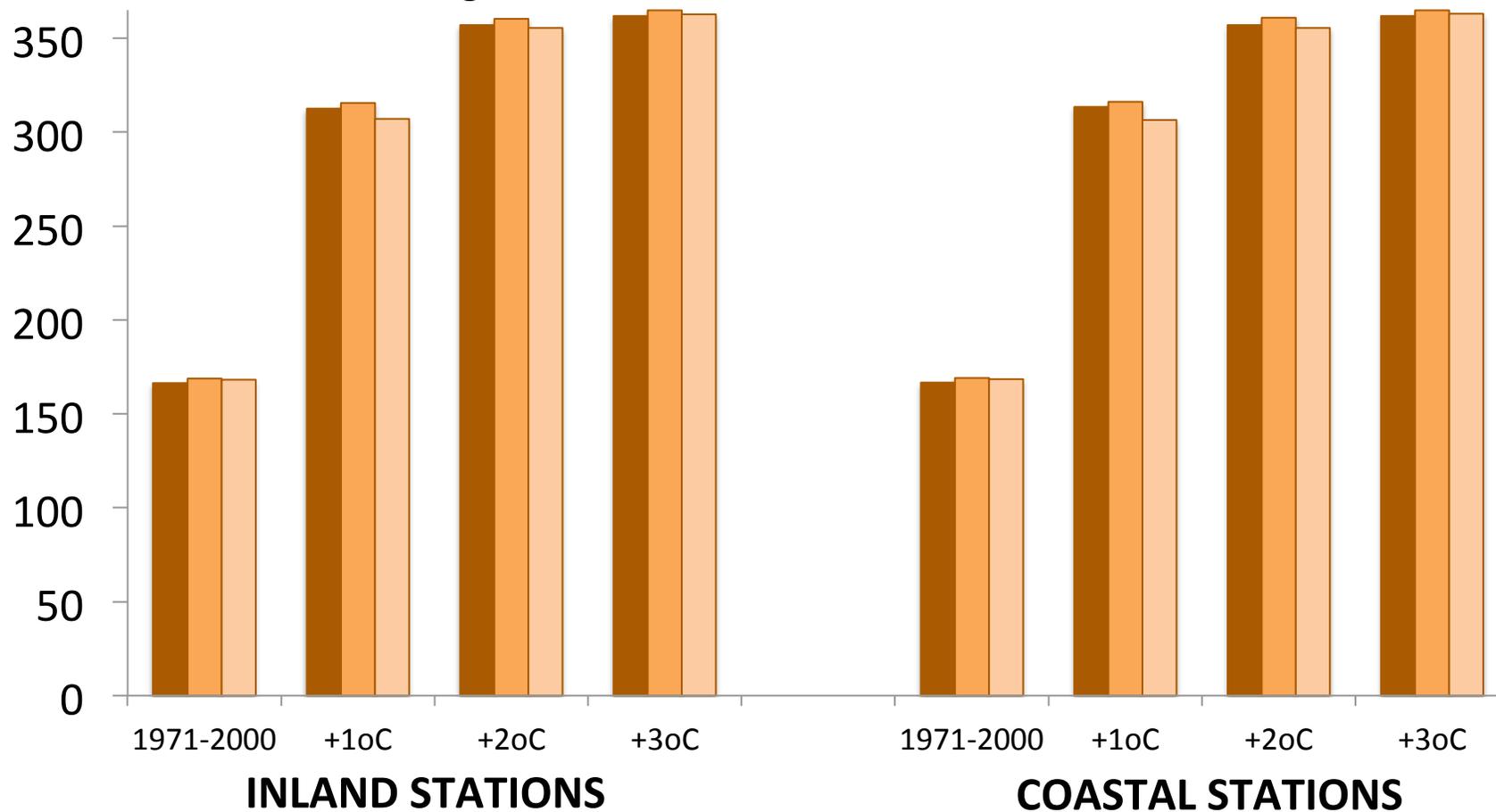
# Annual Cumulative Cooling Degree-Days 85oF Threshold

Average of all **Good** **Fair &** **Poor** CMIP3 Models

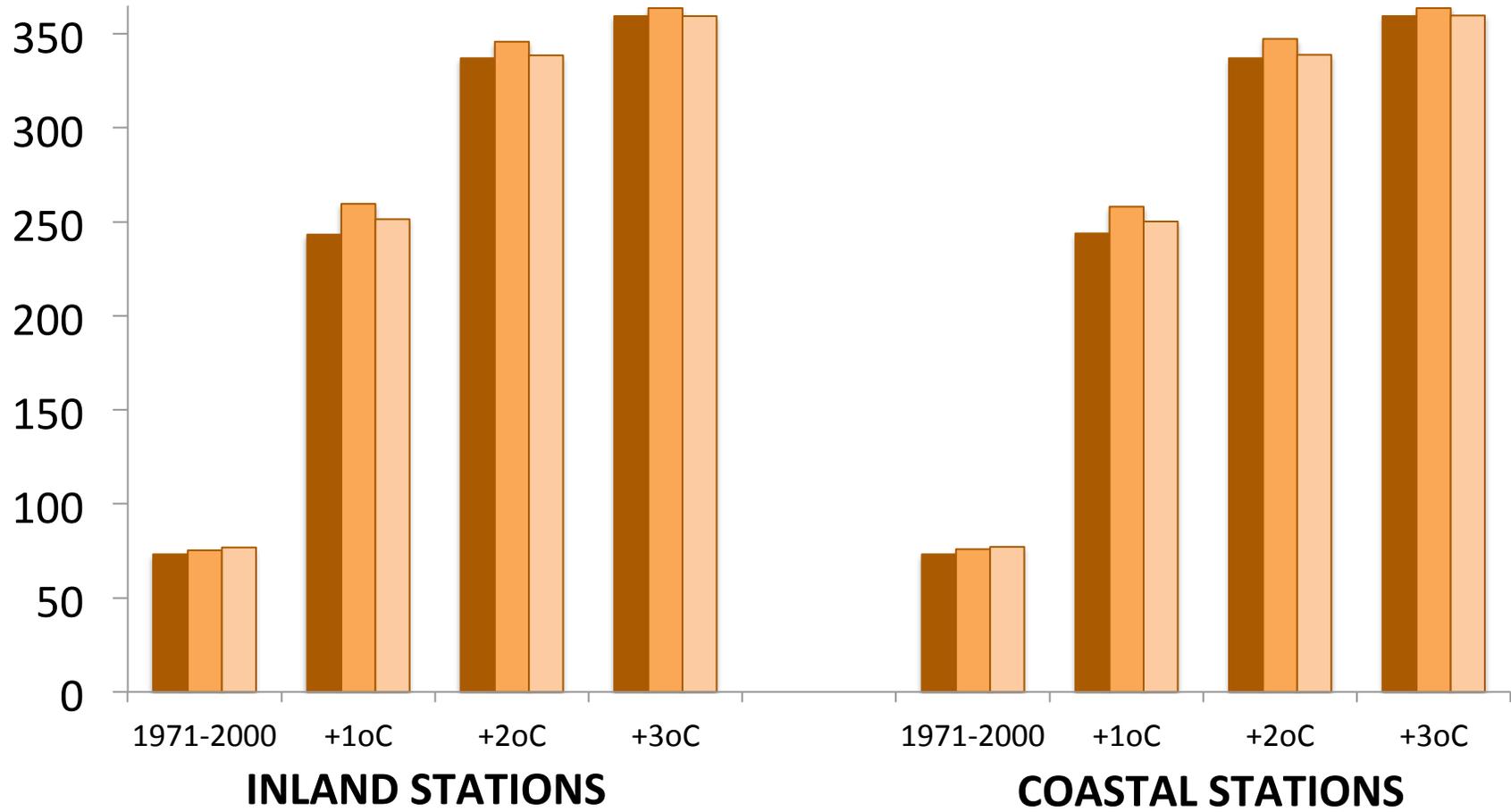


# Days per Year with Daily Maximum Temperature > Historical 50th Quantile or Median (1 in 2 Warmest Days)

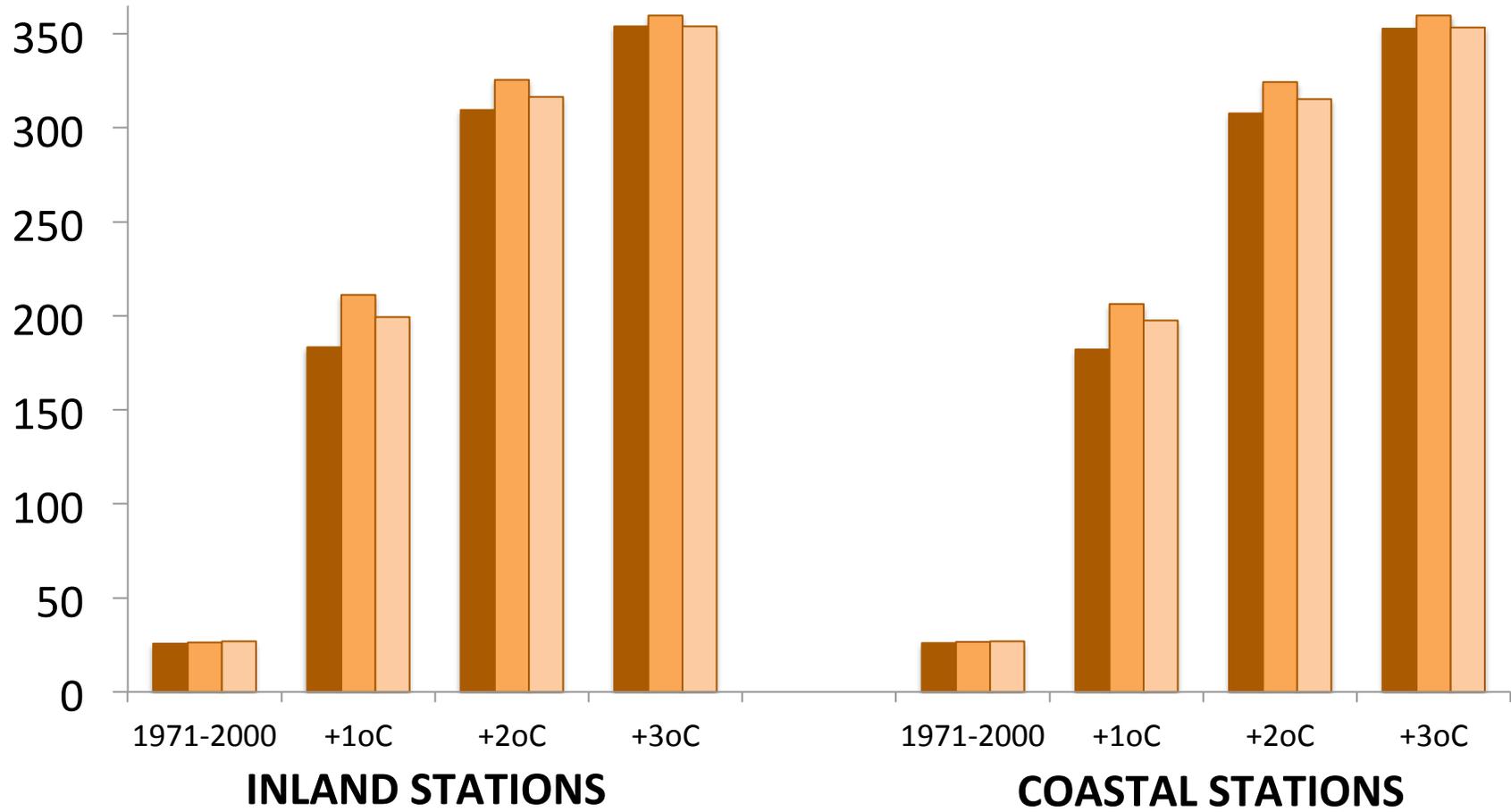
Average of all **Good** **Fair &** **Poor** CMIP3 Models



**Days per Year with Daily Maximum Temperature >  
Historical 75th Quantile (1 in 4 Warmest Days)**  
Average of all **Good** **Fair &** **Poor** CMIP3 Models

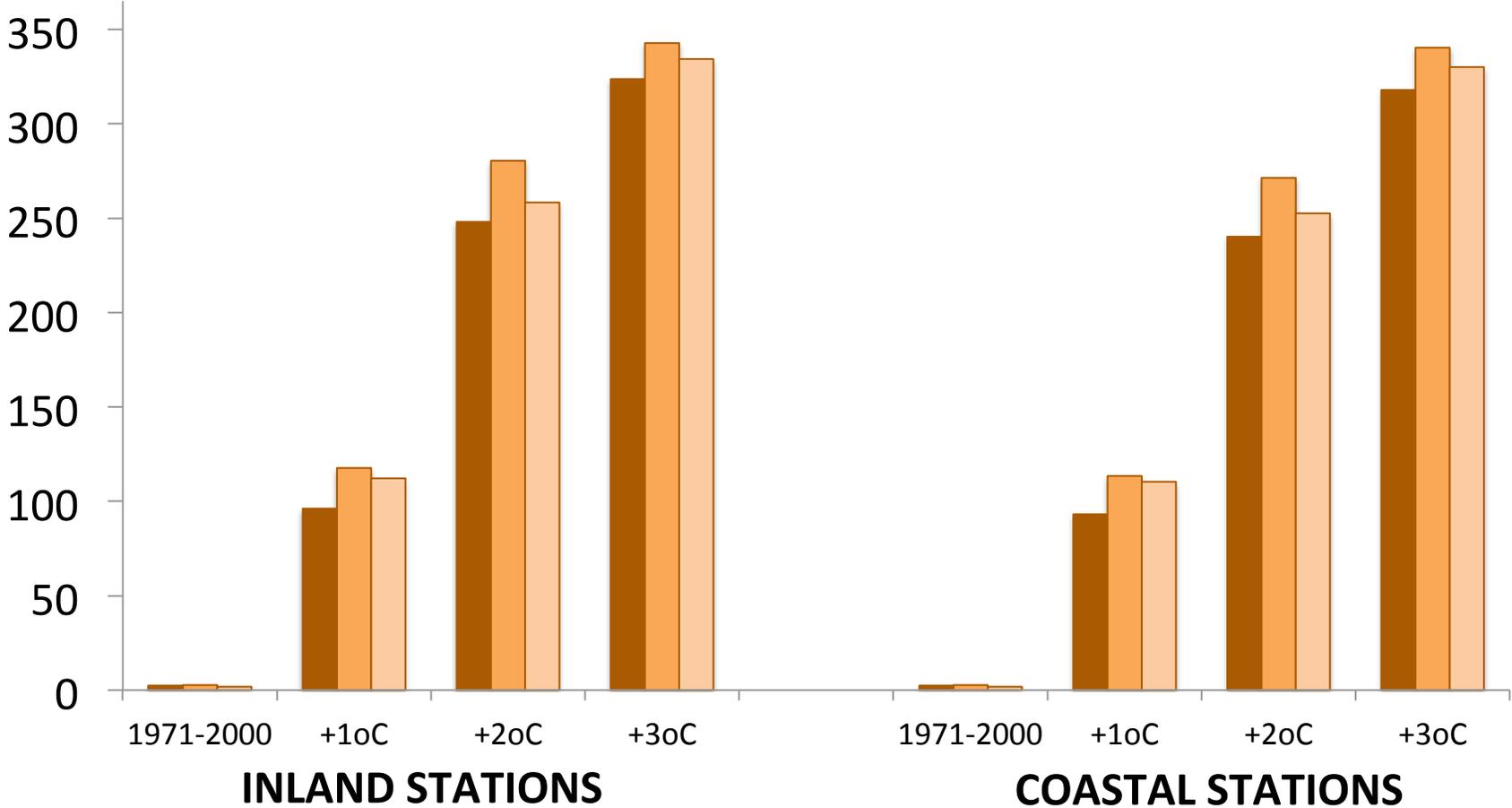


**Days per Year with Daily Maximum Temperature >  
Historical 90th Quantile (1 in 10 Warmest Days)**  
Average of all **Good** **Fair &** **Poor** CMIP3 Models



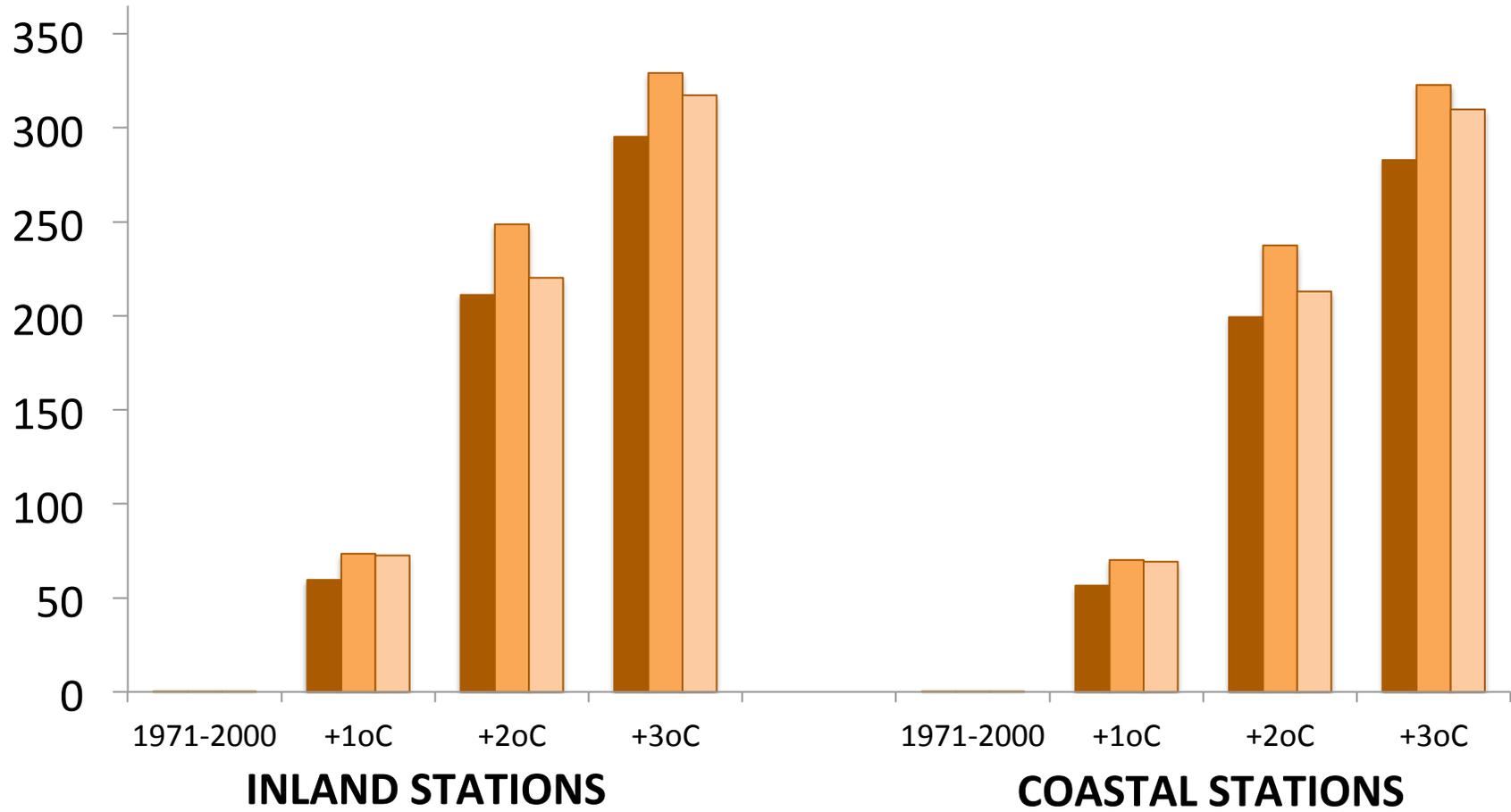
# Days per Year with Daily Maximum Temperature > Historical 99th Quantile (1 in 100 Warmest Days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models

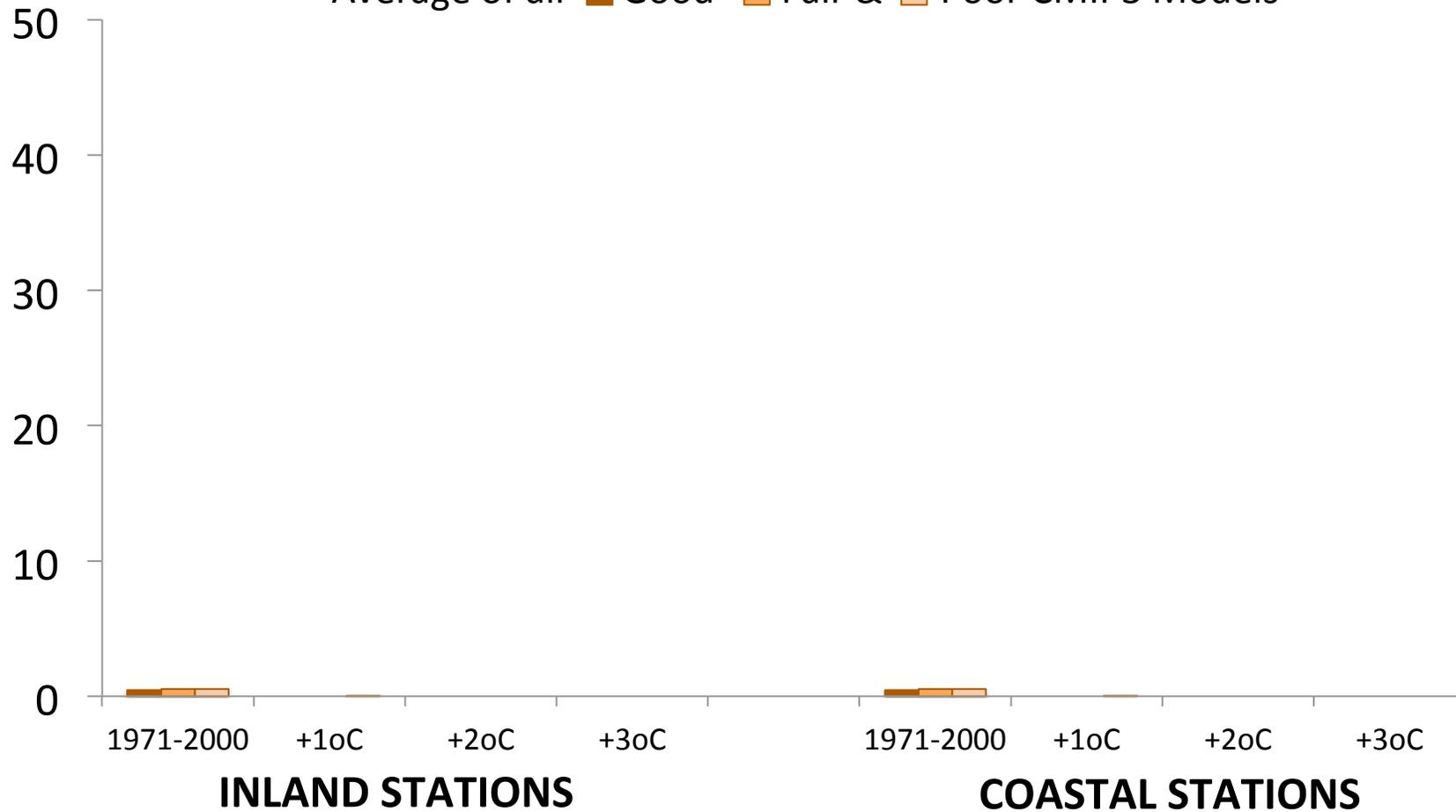


# Days per Year with Daily Maximum Temperature > Historical 99.9th Quantile (1 in 1000 Warmest Days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models

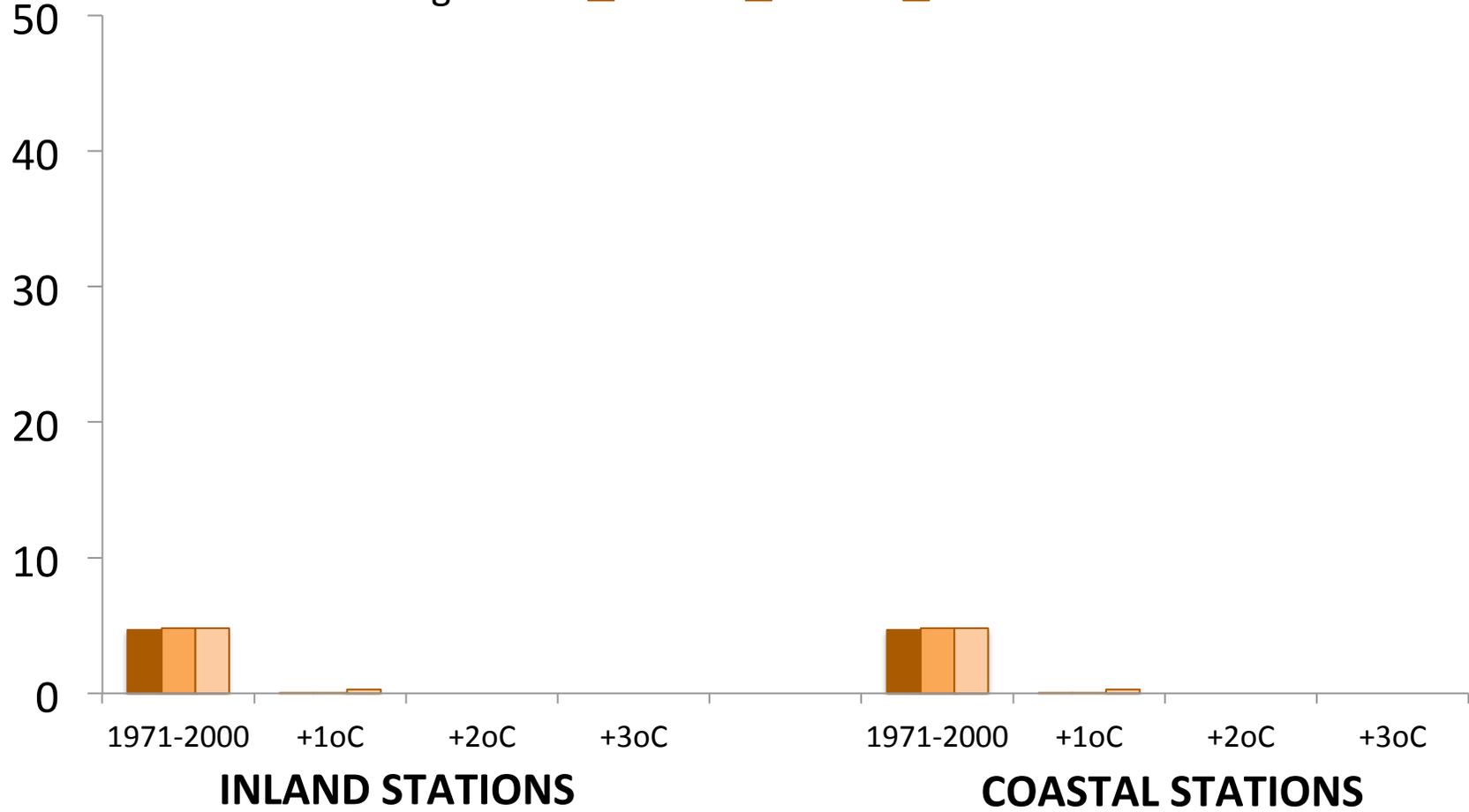


**Days per Year with Nighttime Minimum Temperature <  
Historical 0.1th Quantile (1 in 1000 Coolest Days)**  
Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models

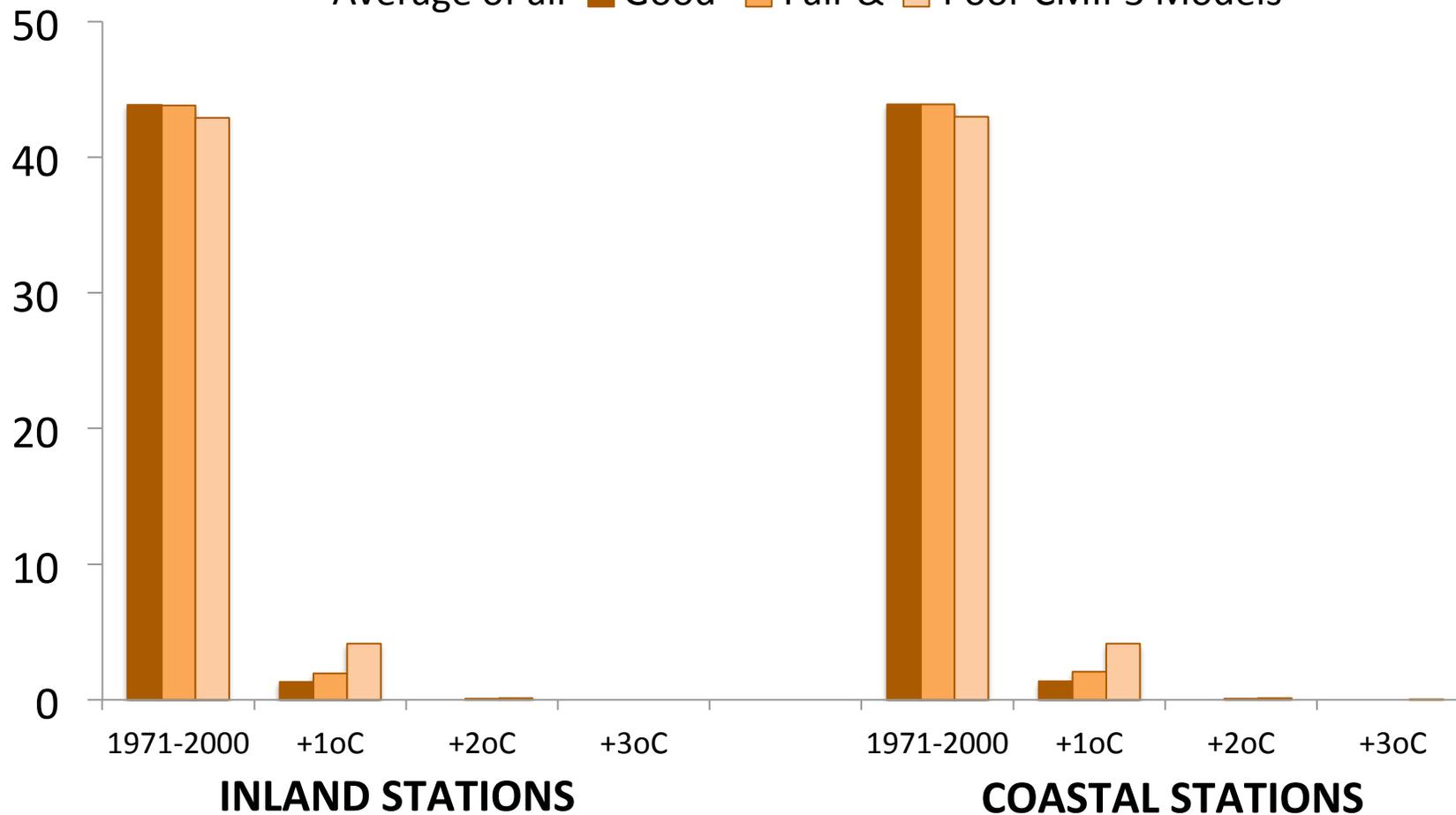


# Days per Year with Nighttime Minimum Temperature < Historical 1st Quantile (1 in 100 Coolest Days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models

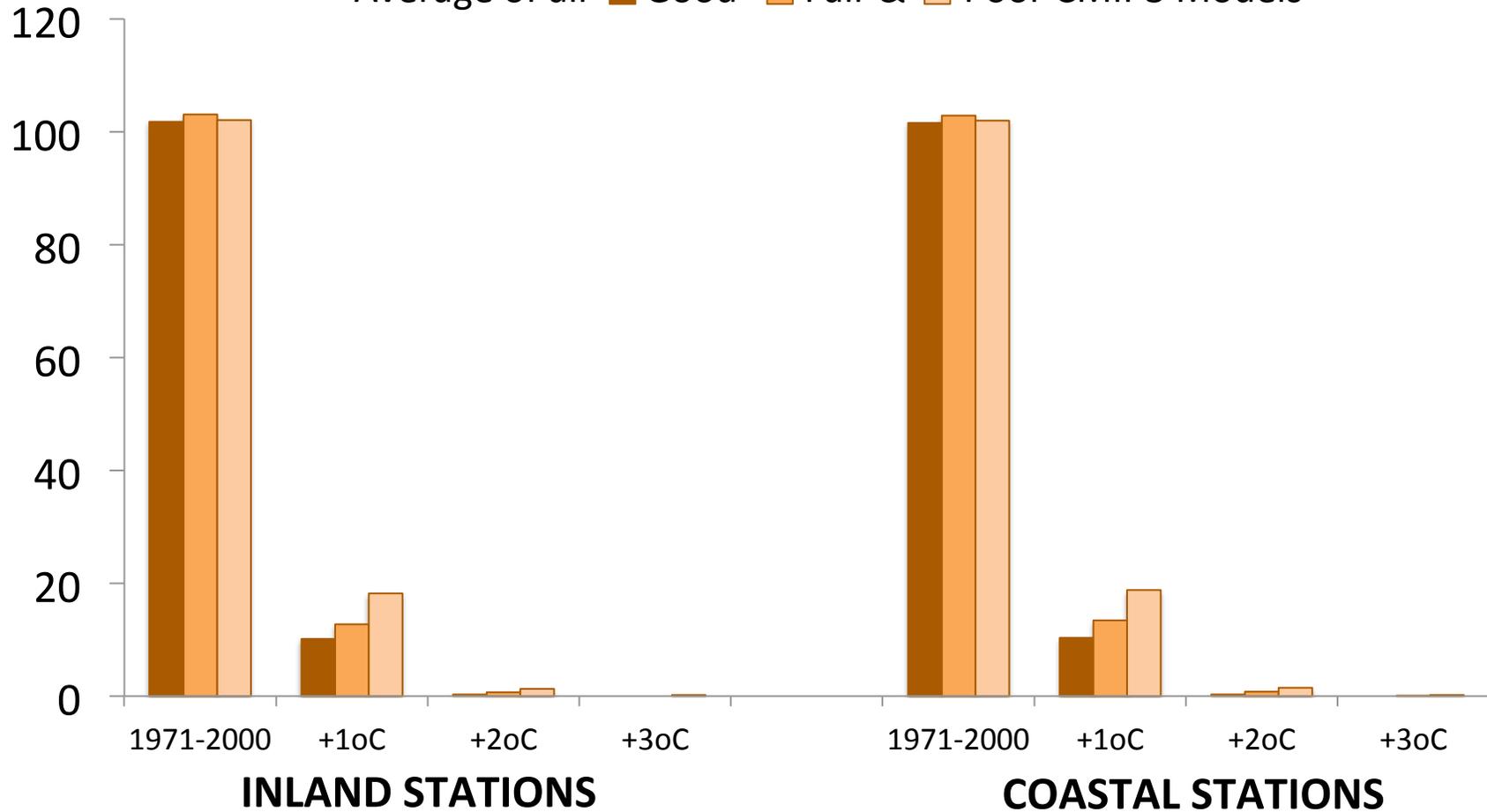


**Days per Year with Nighttime Minimum Temperature <  
Historical 10th Quantile (1 in 10 Coolest Days)**  
Average of all **Good** **Fair &** **Poor** CMIP3 Models



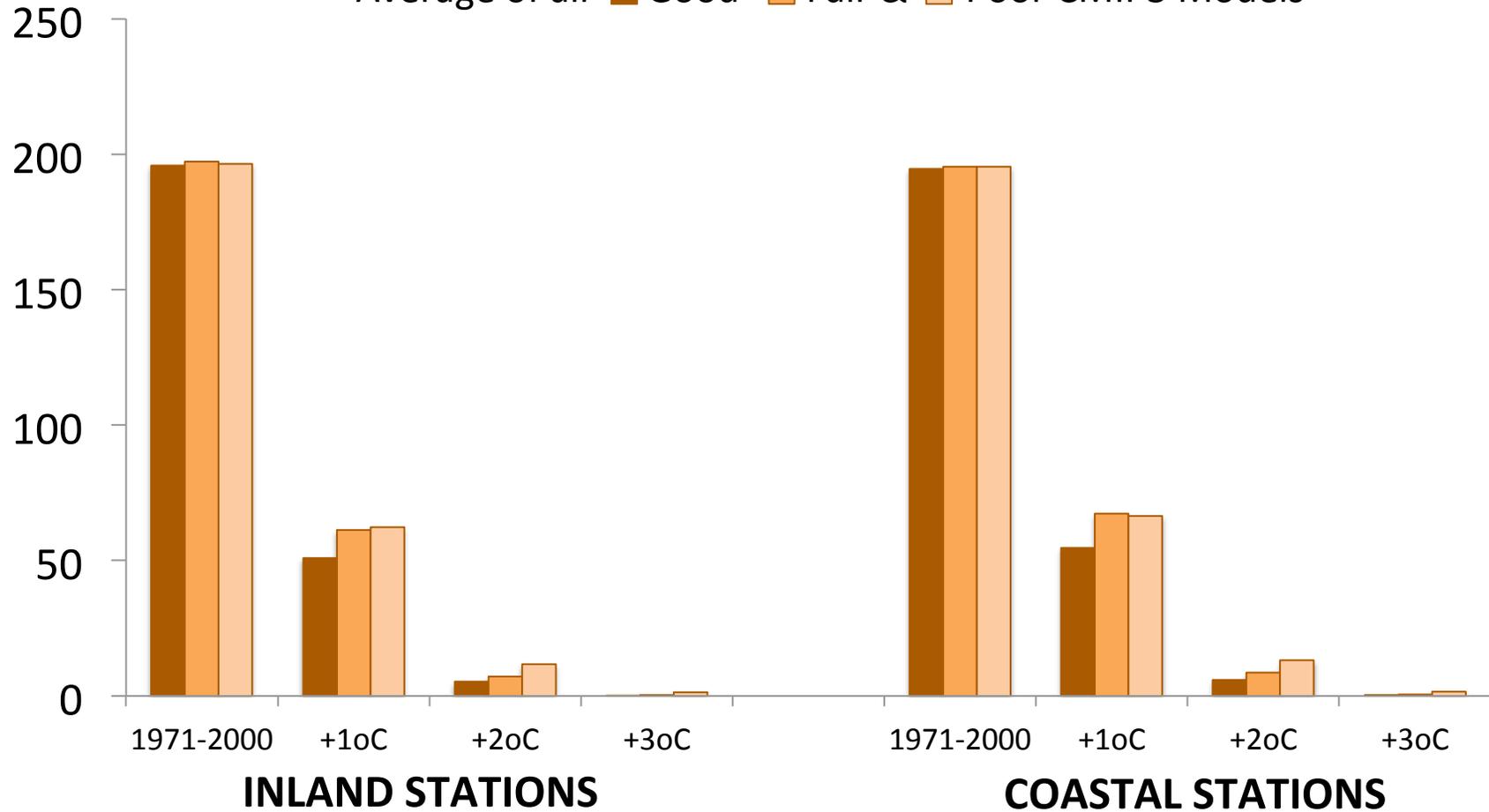
# Days per Year with Nighttime Minimum Temperature < Historical 25th Quantile (1 in 4 Coolest Days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



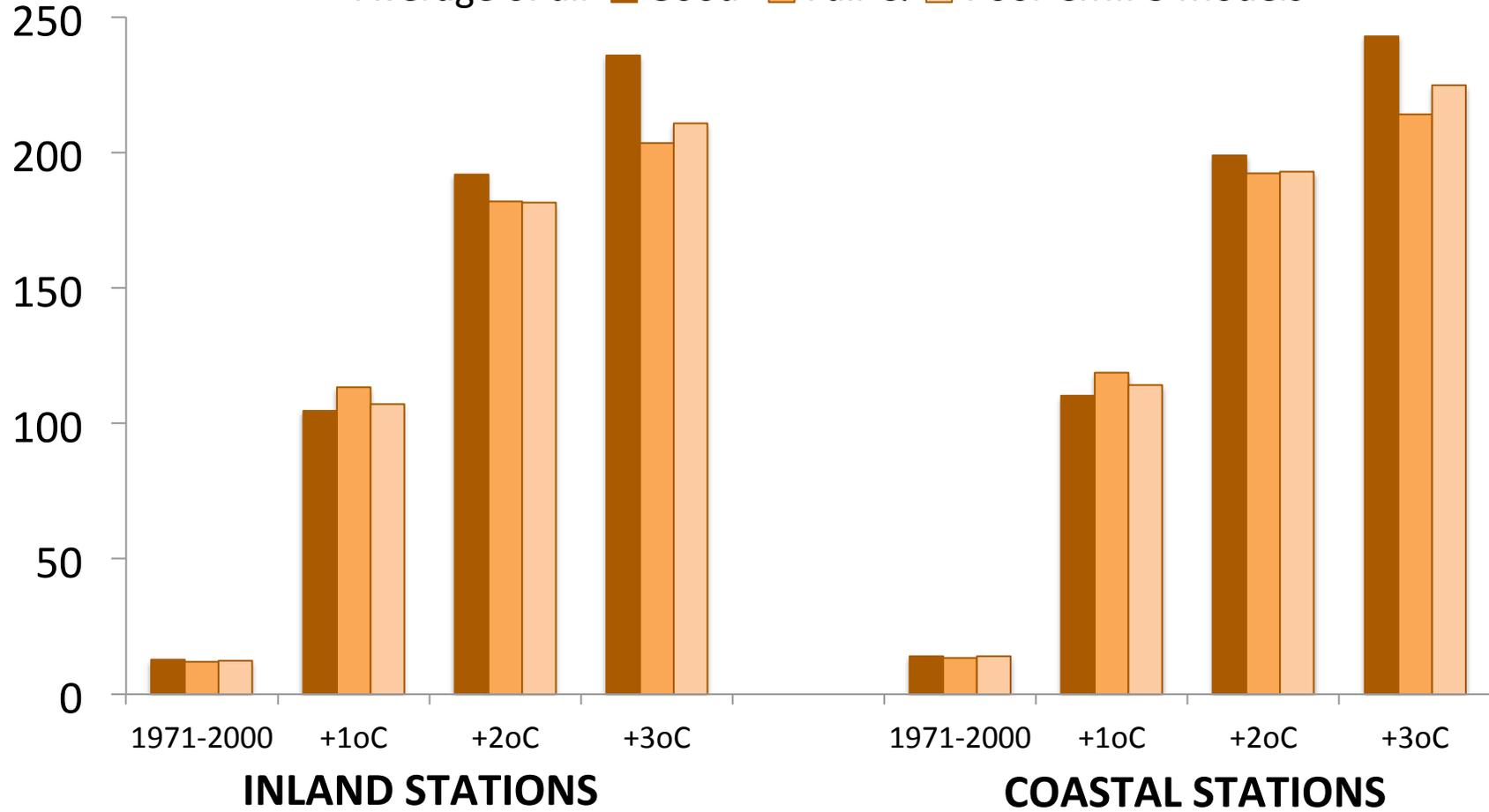
# Days per Year with Nighttime Minimum Temperature < Historical 50th Quantile or Median (1 in 2 Coolest Days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



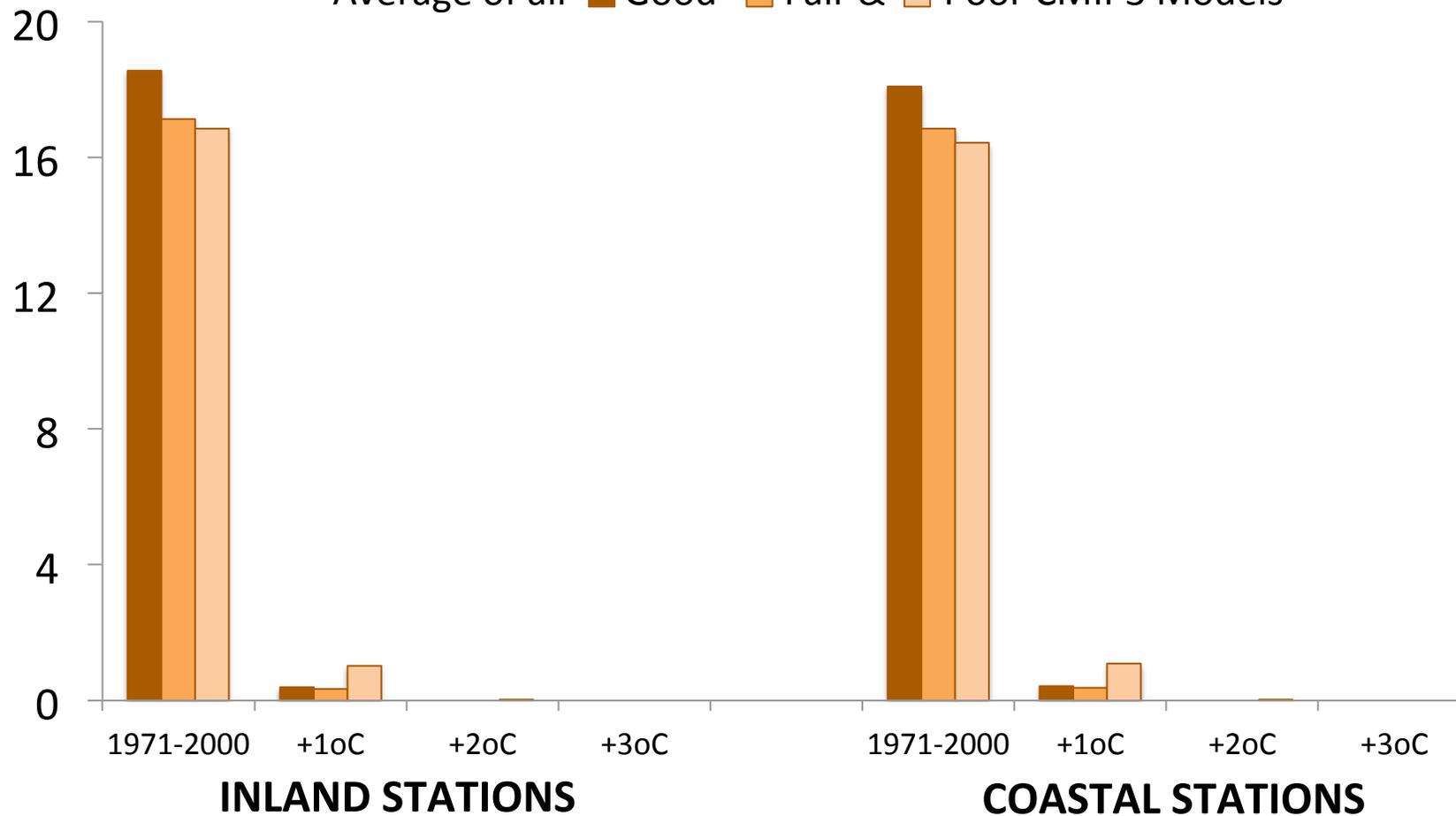
**Hot/Dry Days: average number of days per year with no rain and daily maximum temperature above 1-in-10 warmest historical day**

Average of all **Good** **Fair &** **Poor** CMIP3 Models



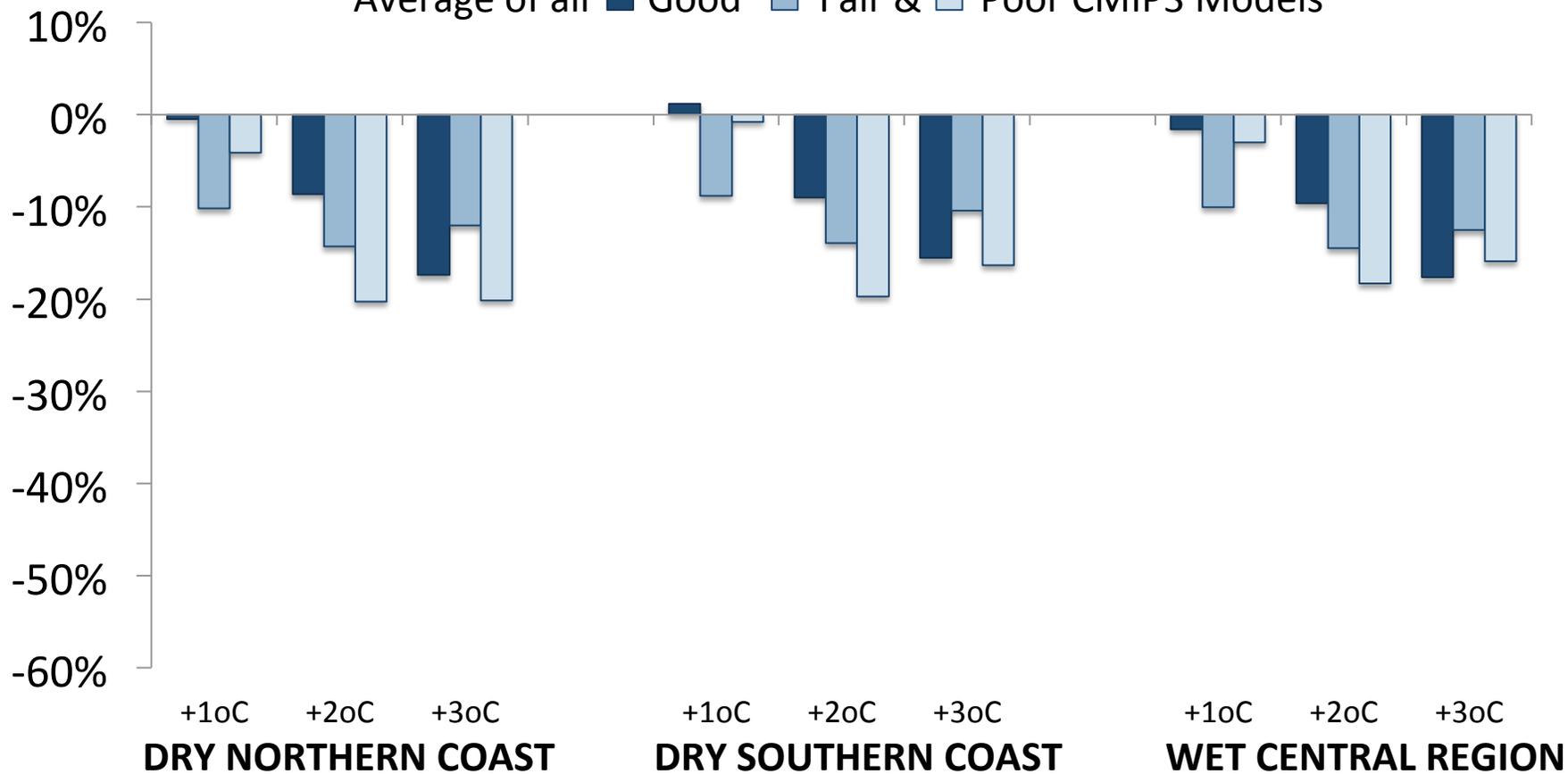
**Cool/Wet Days: average number of rainy days per year with daily maximum temperature below 1-in-10 coolest historical day**

Average of all **Good** **Fair &** **Poor** CMIP3 Models



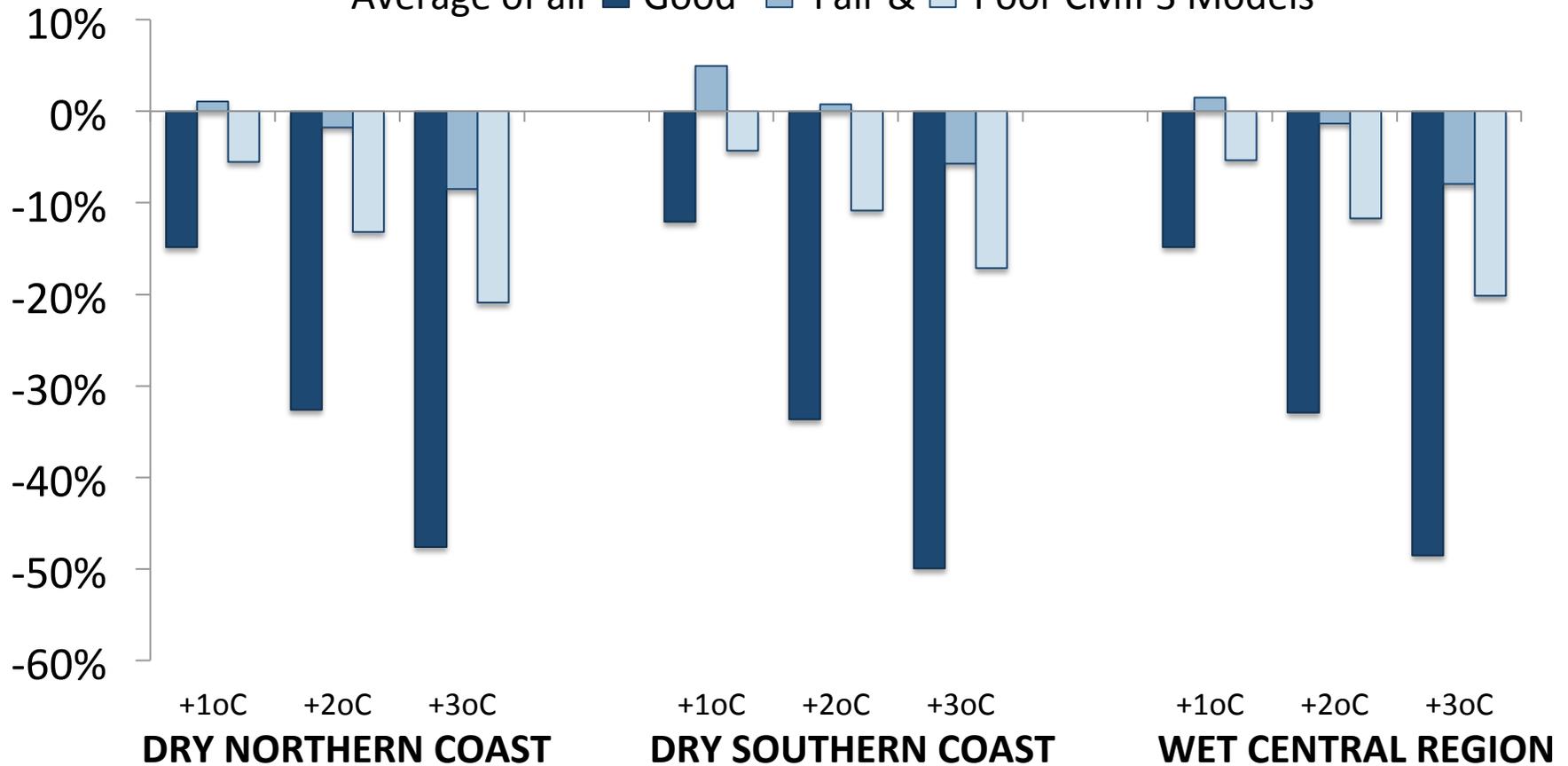
# Precipitation in Dry Season (percentage change)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



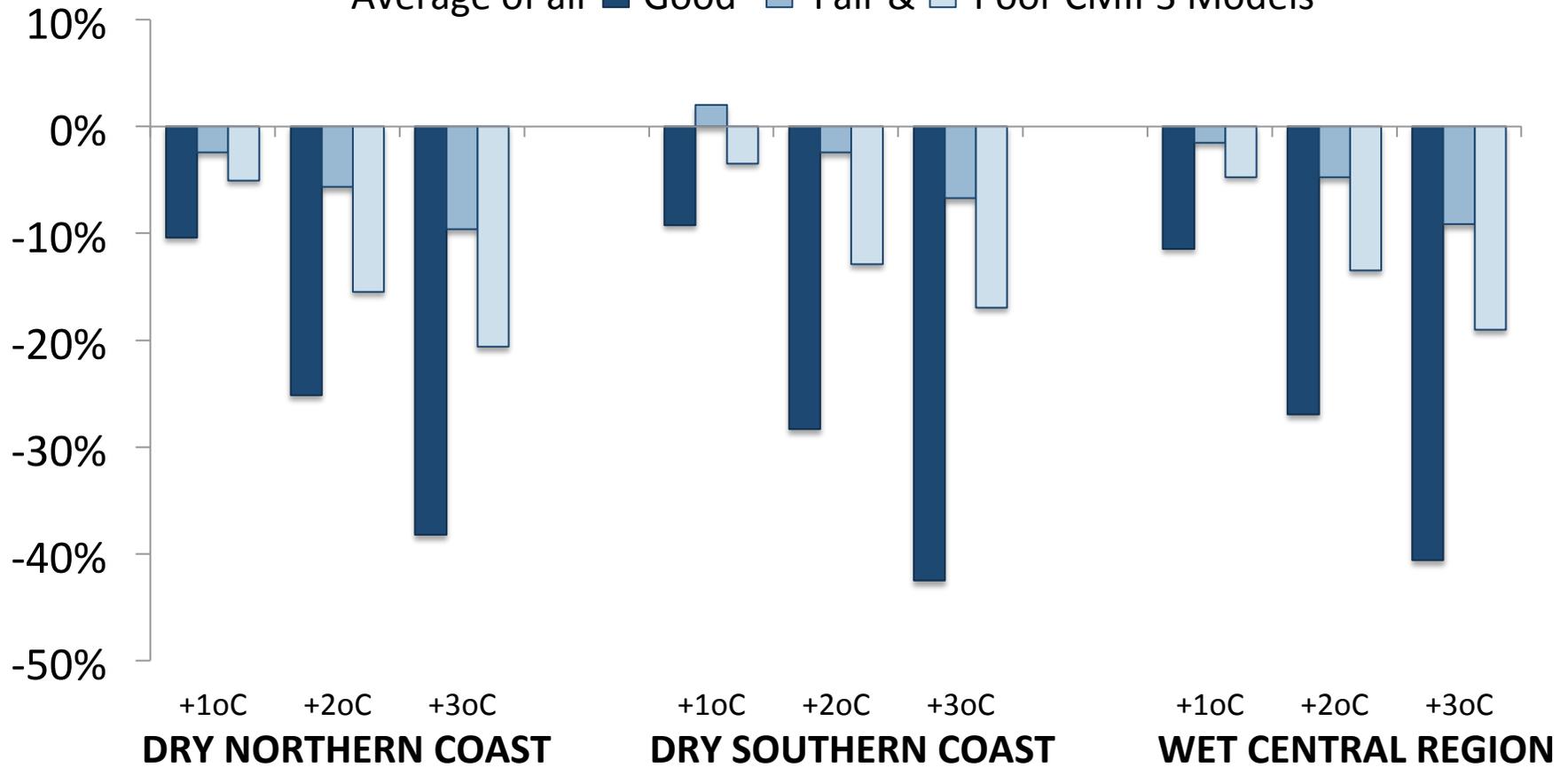
# Precipitation in Wet Season (percentage change)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



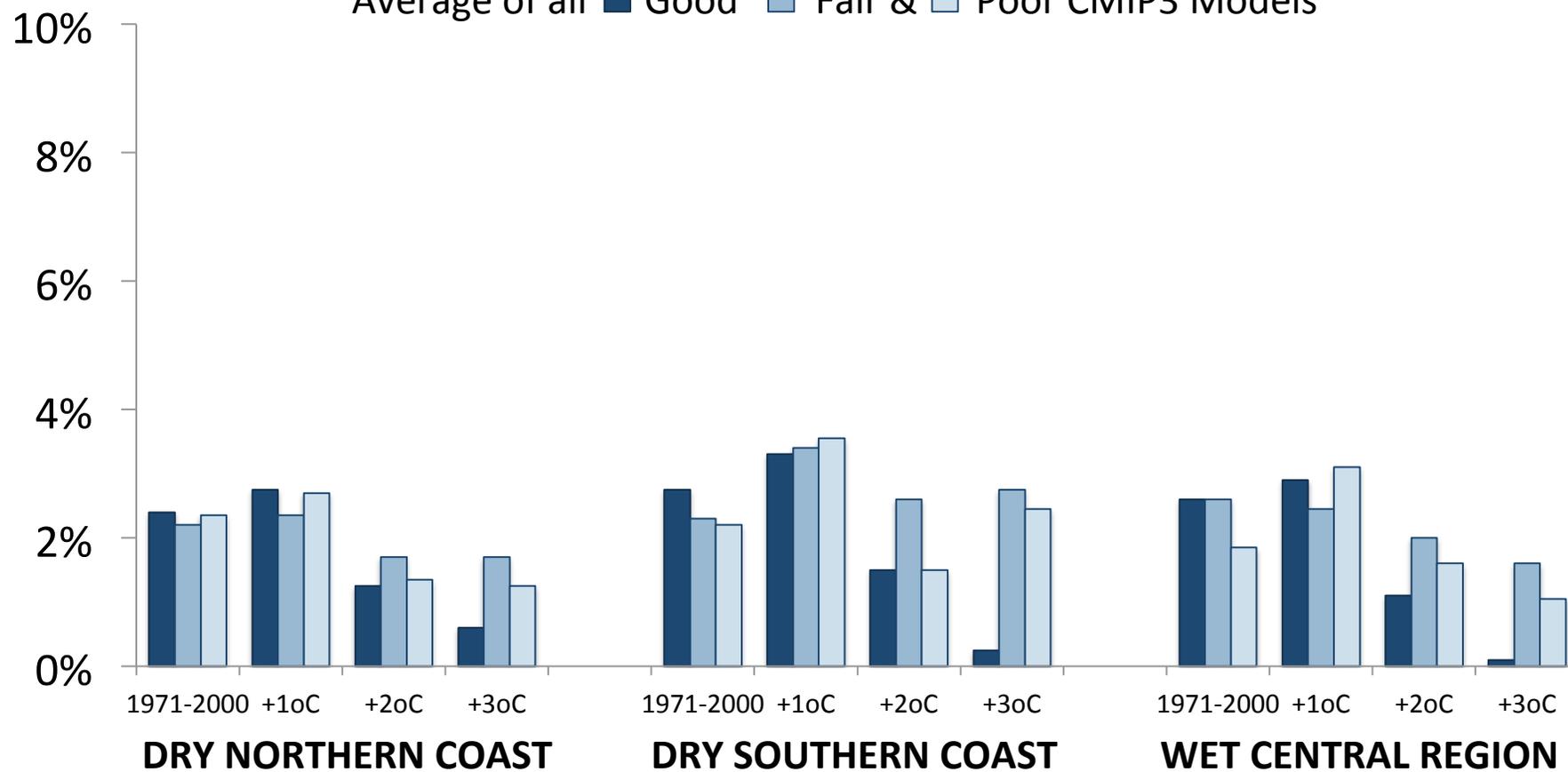
# Annual Precipitation (percentage change)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



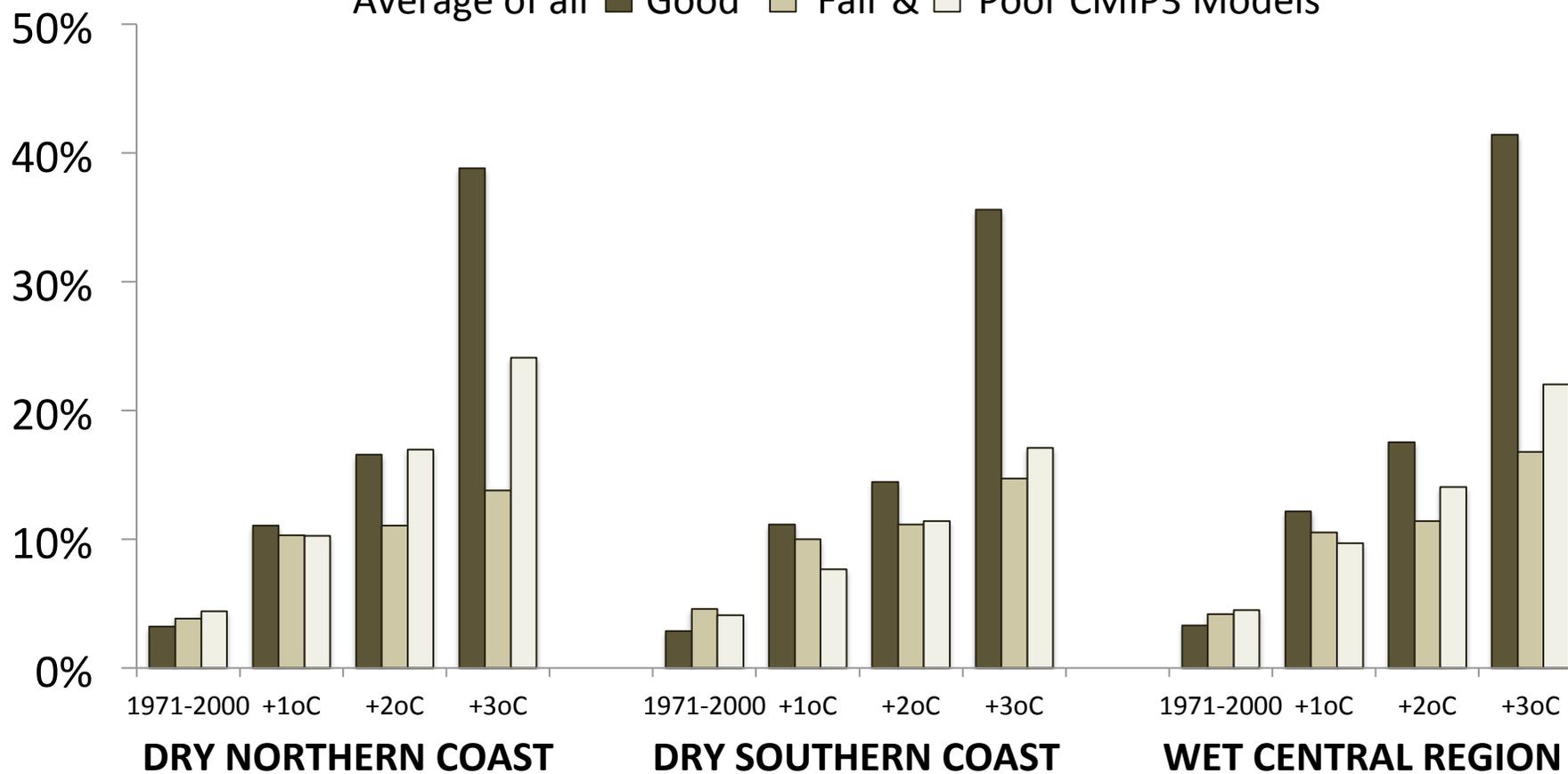
# Wet Years (percent)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



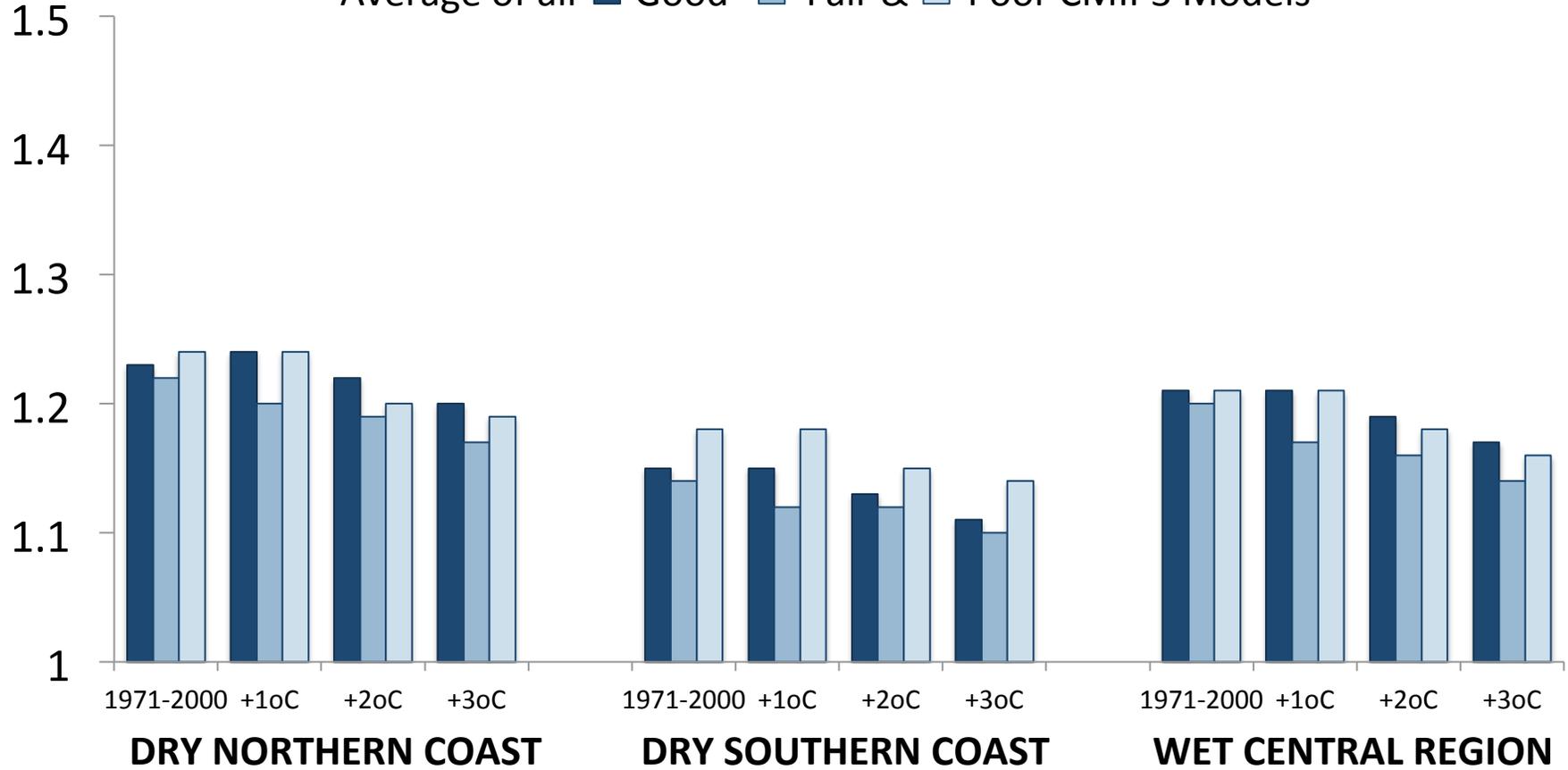
# Dry Years (percent)

Average of all  Good  Fair &  Poor CMIP3 Models



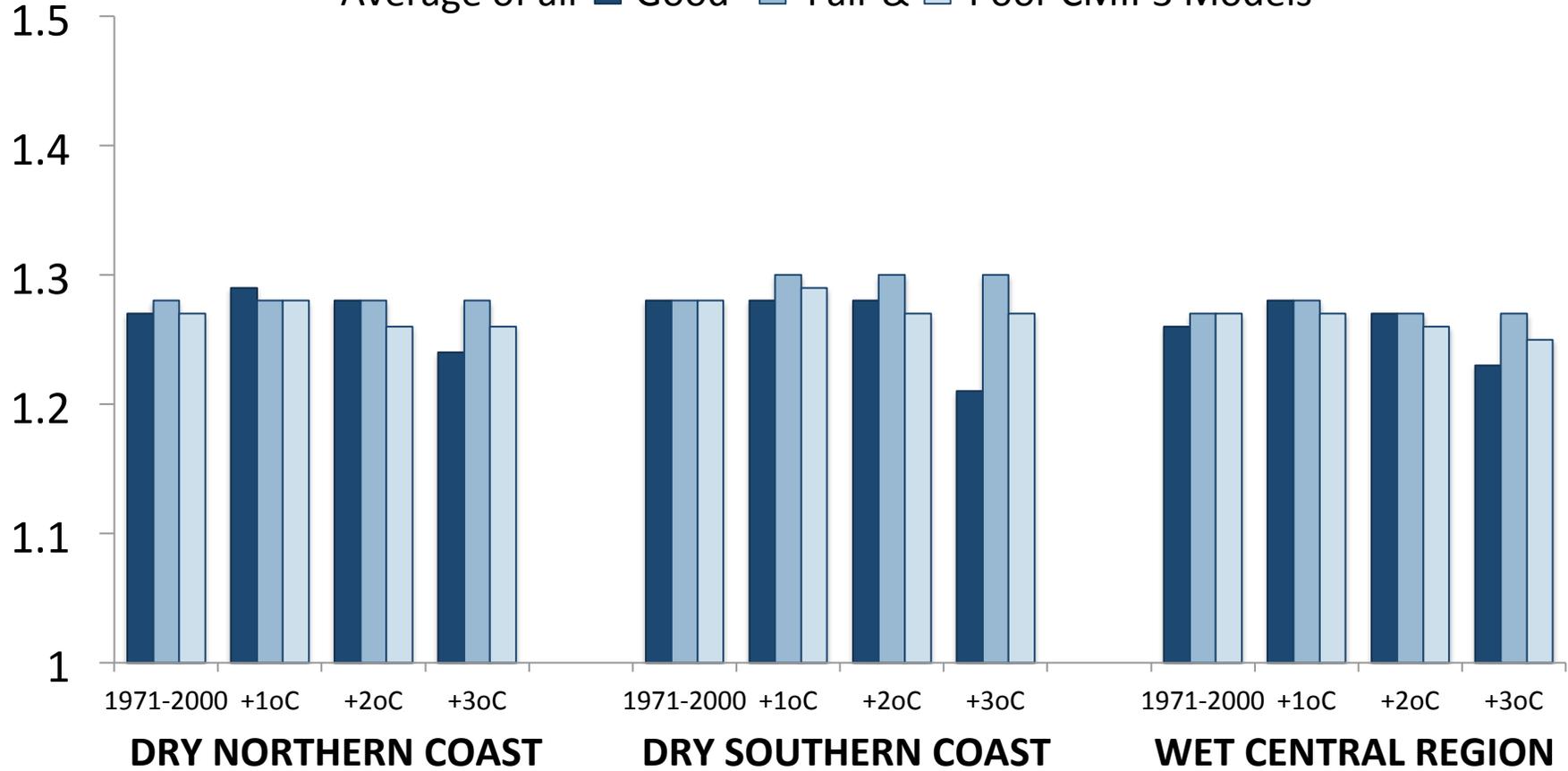
# Standard Deviation of Precipitation in Dry Season (mm)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



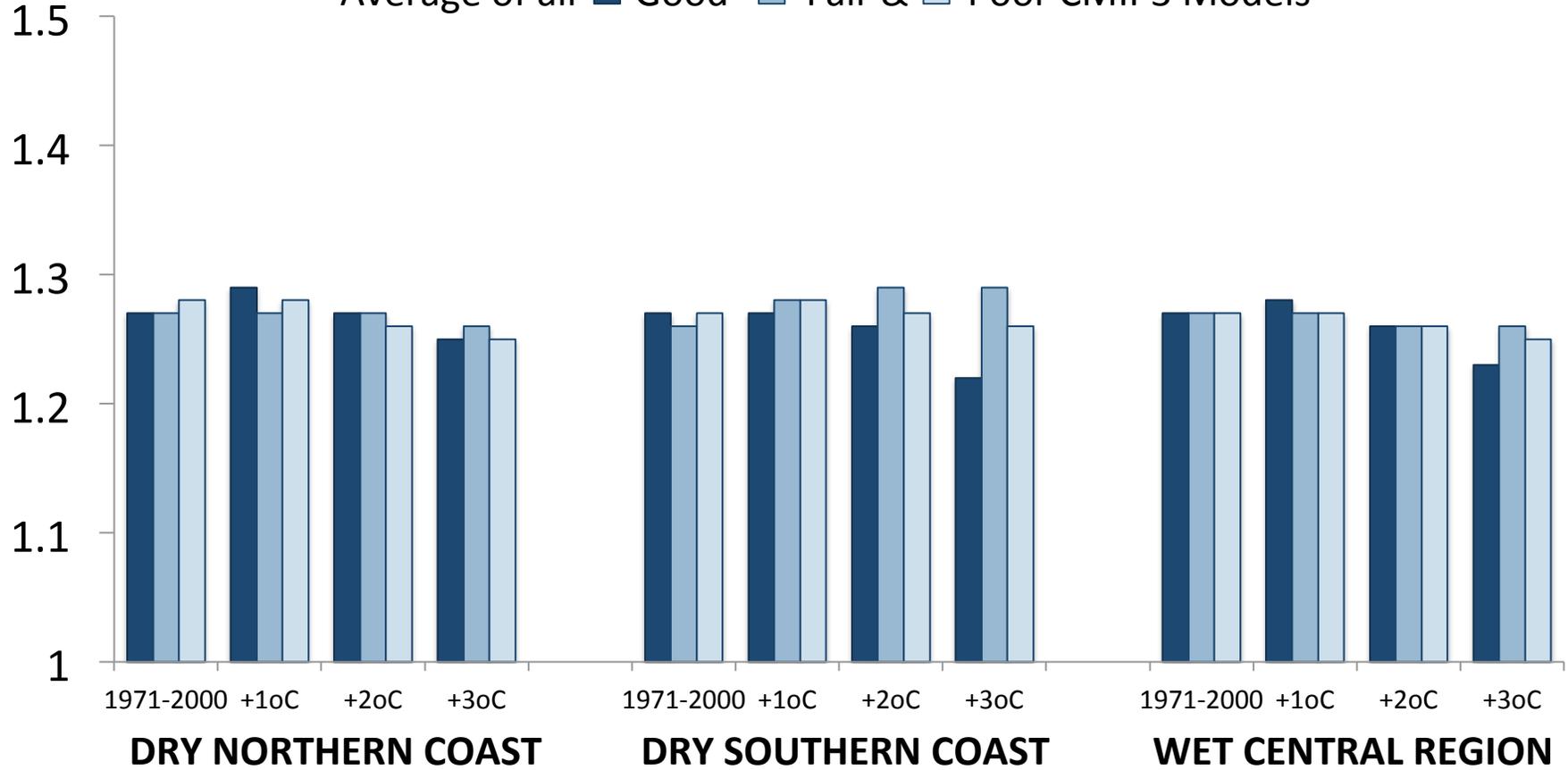
# Standard Deviation of Precipitation in Wet Season (mm)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



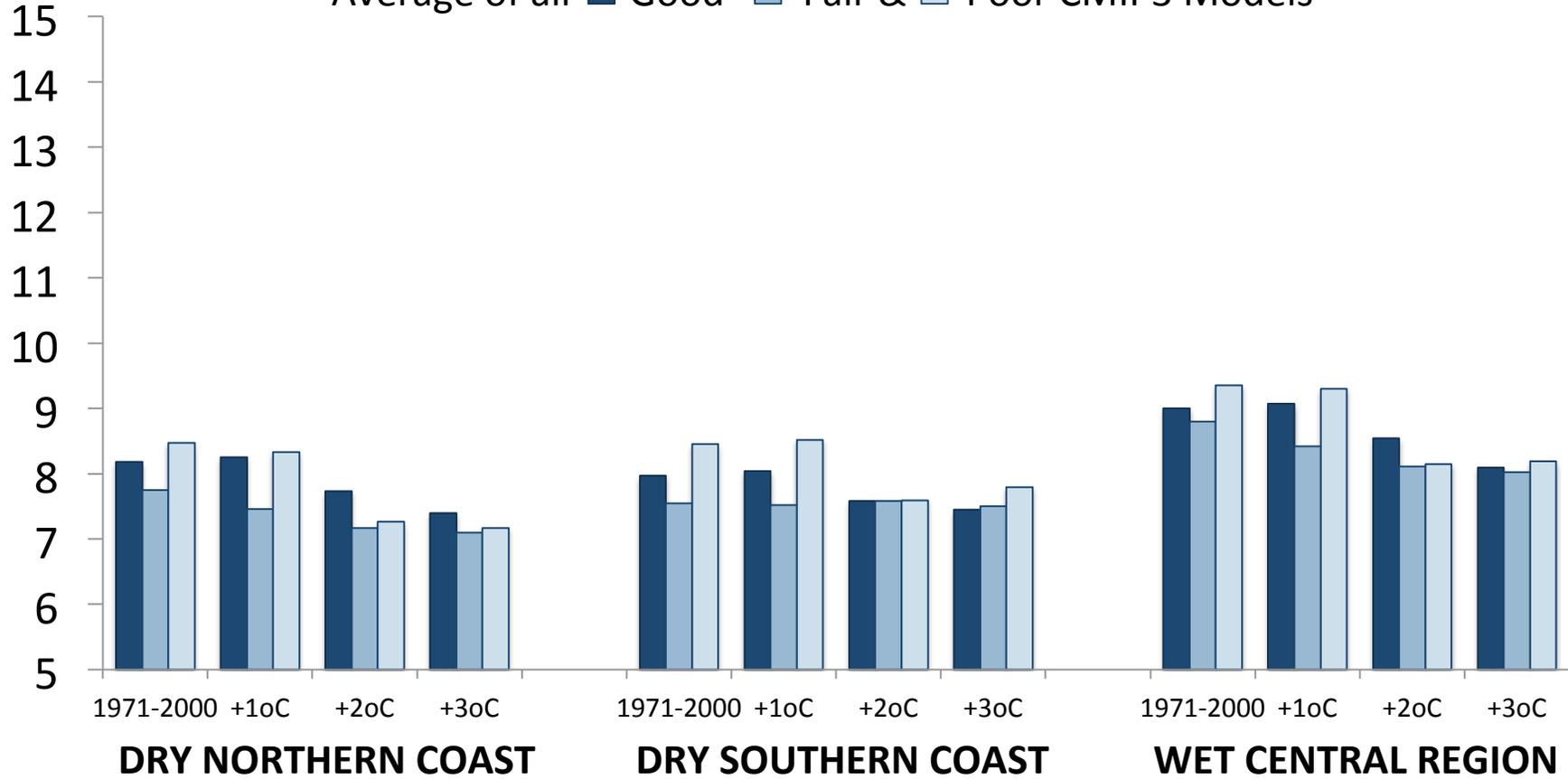
# Standard Deviation of Precipitation Annual (mm)

Average of all **Good** **Fair &** **Poor** CMIP3 Models

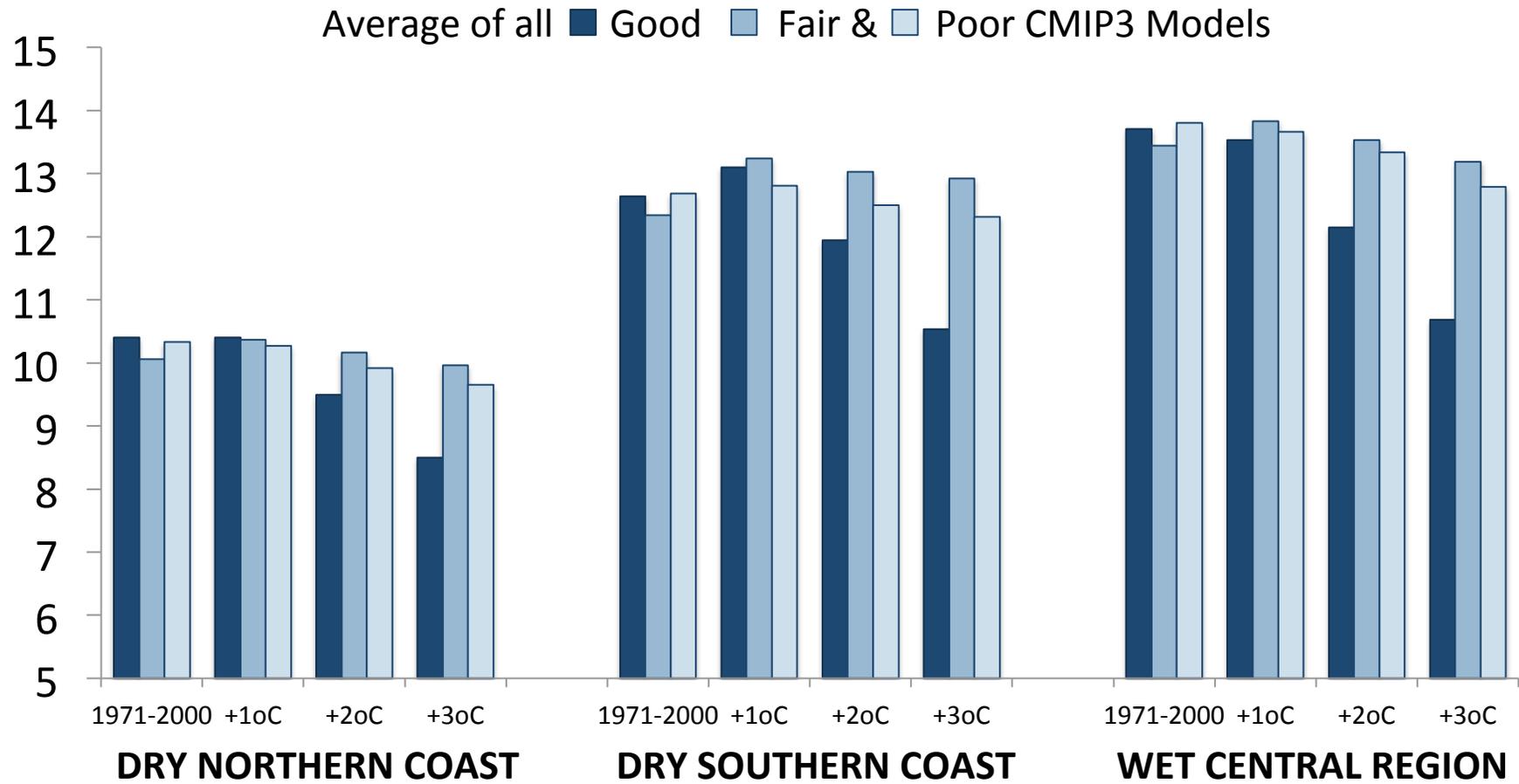


# Precipitation Intensity in Dry Season (mm/day)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models

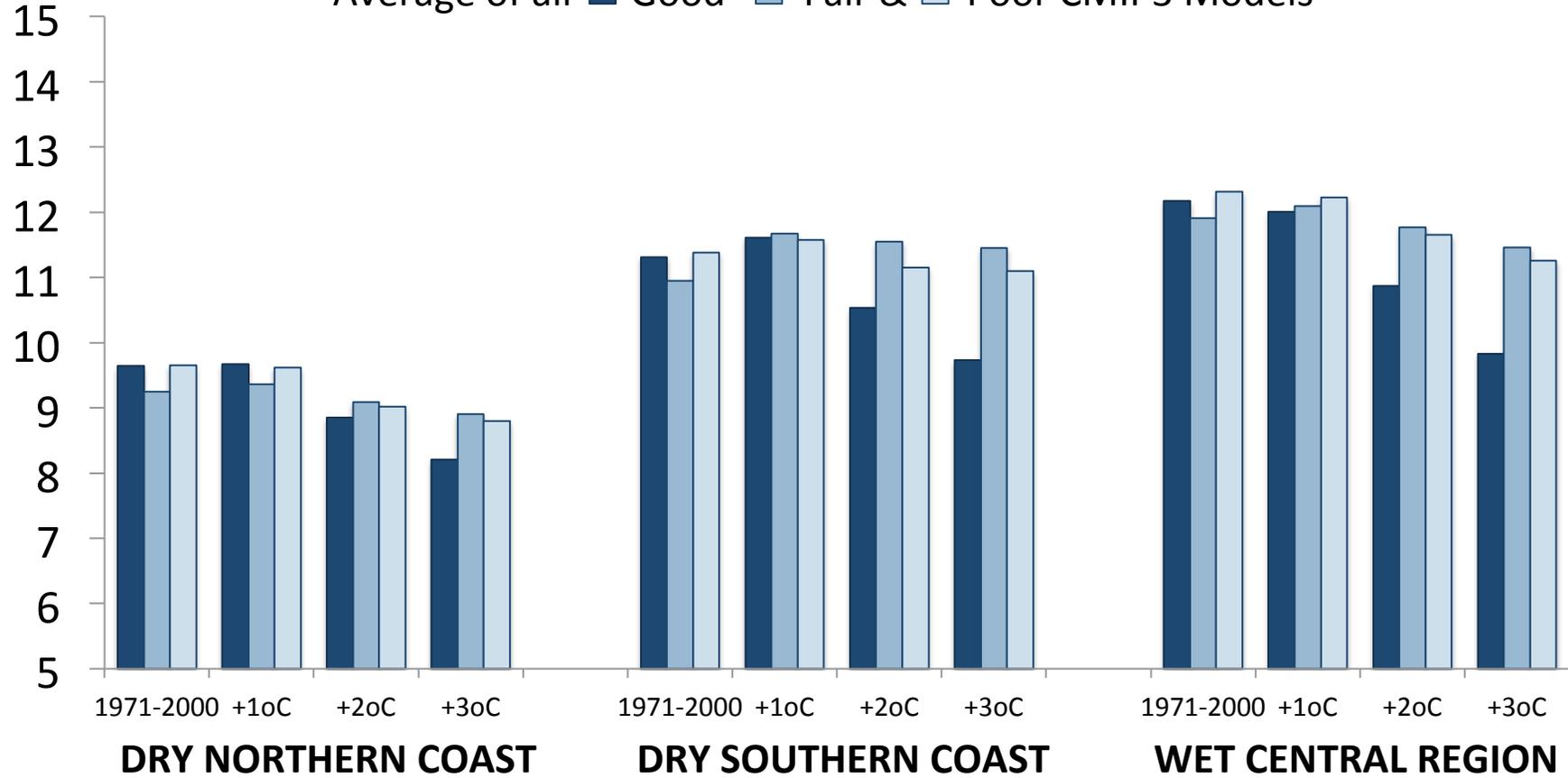


# Precipitation Intensity in Wet Season (mm/day)



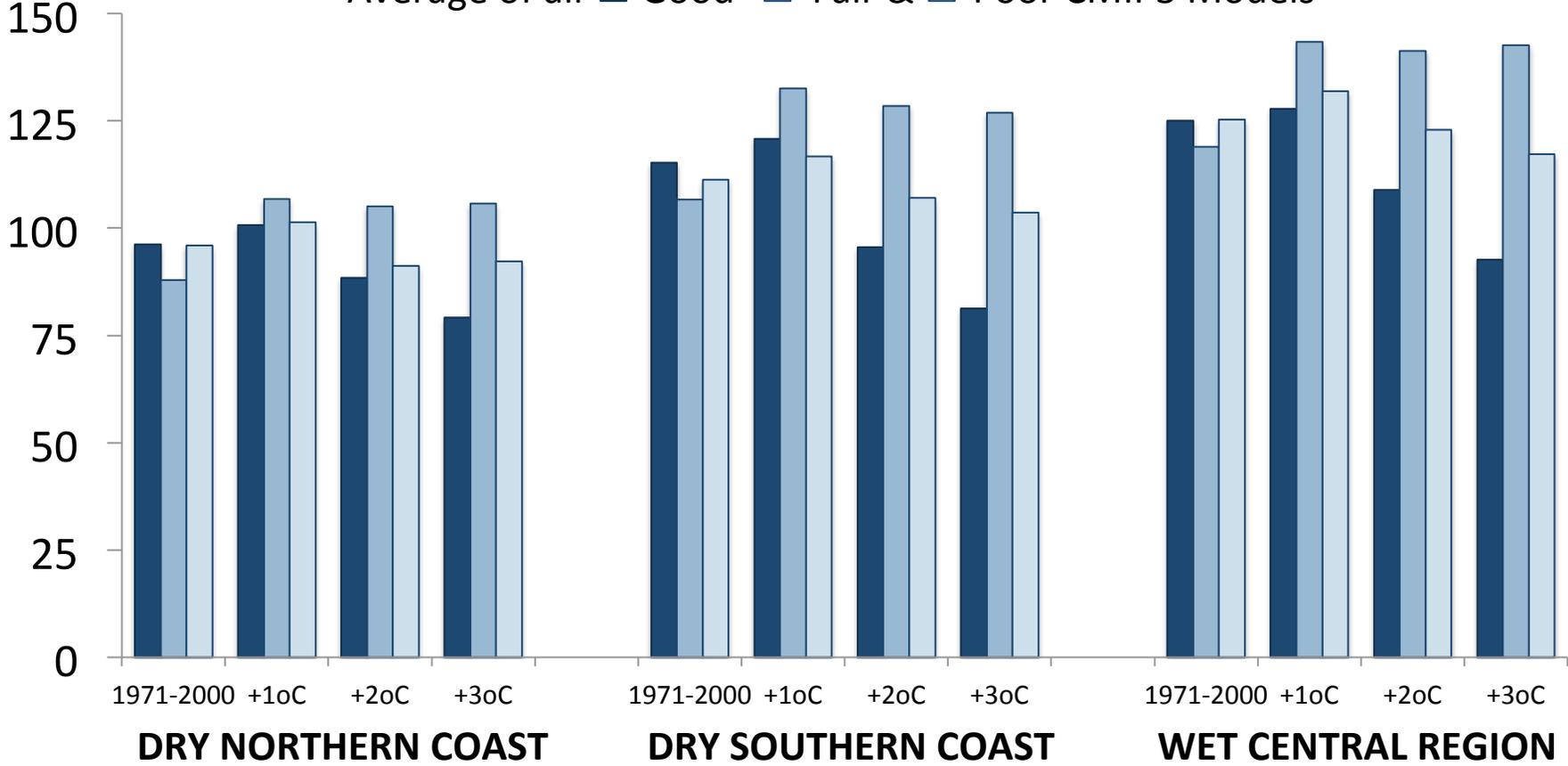
# Annual Precipitation Intensity (mm/day)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



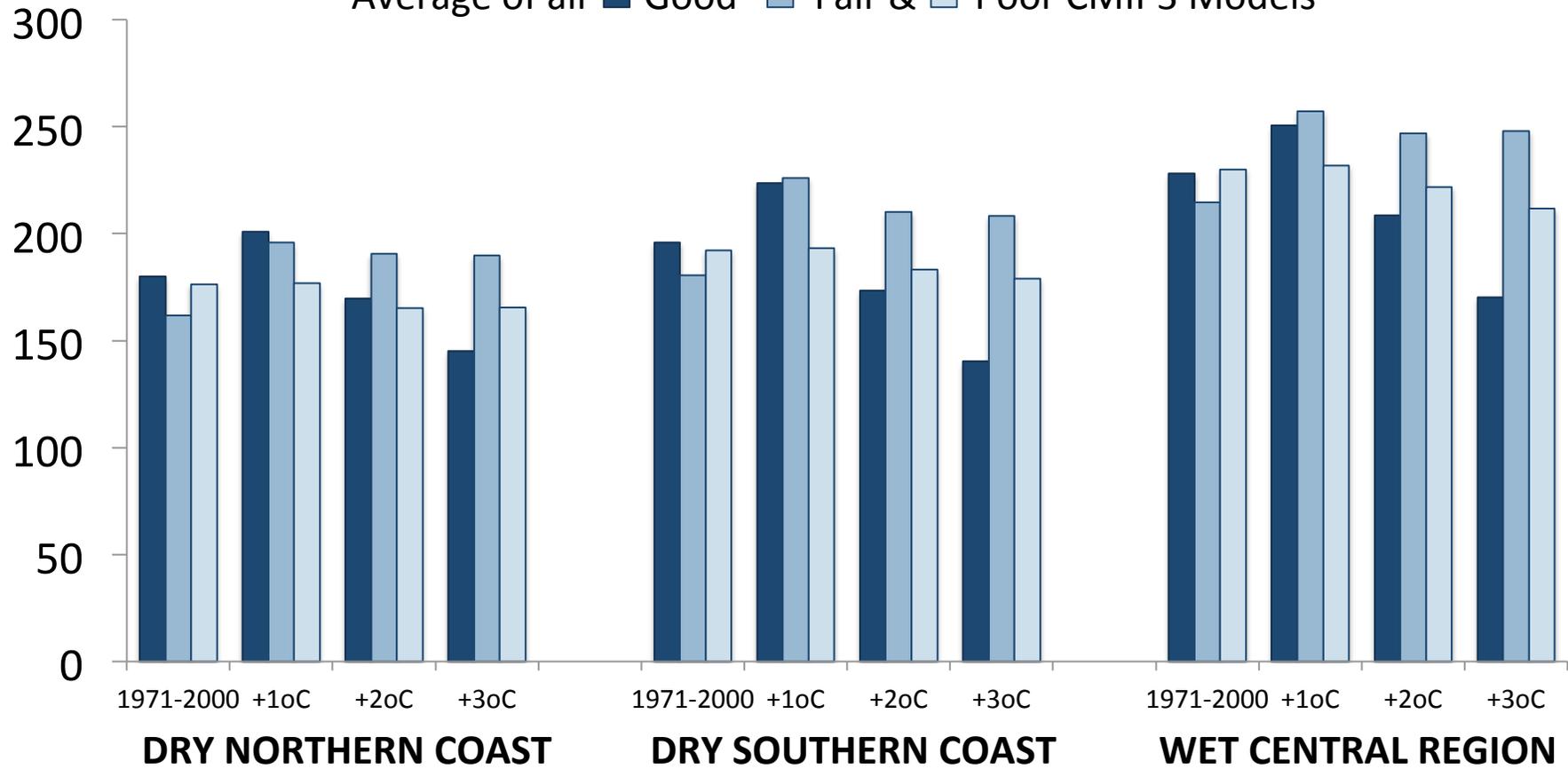
# Precipitation in Wettest 1 Day of the Year (mm)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



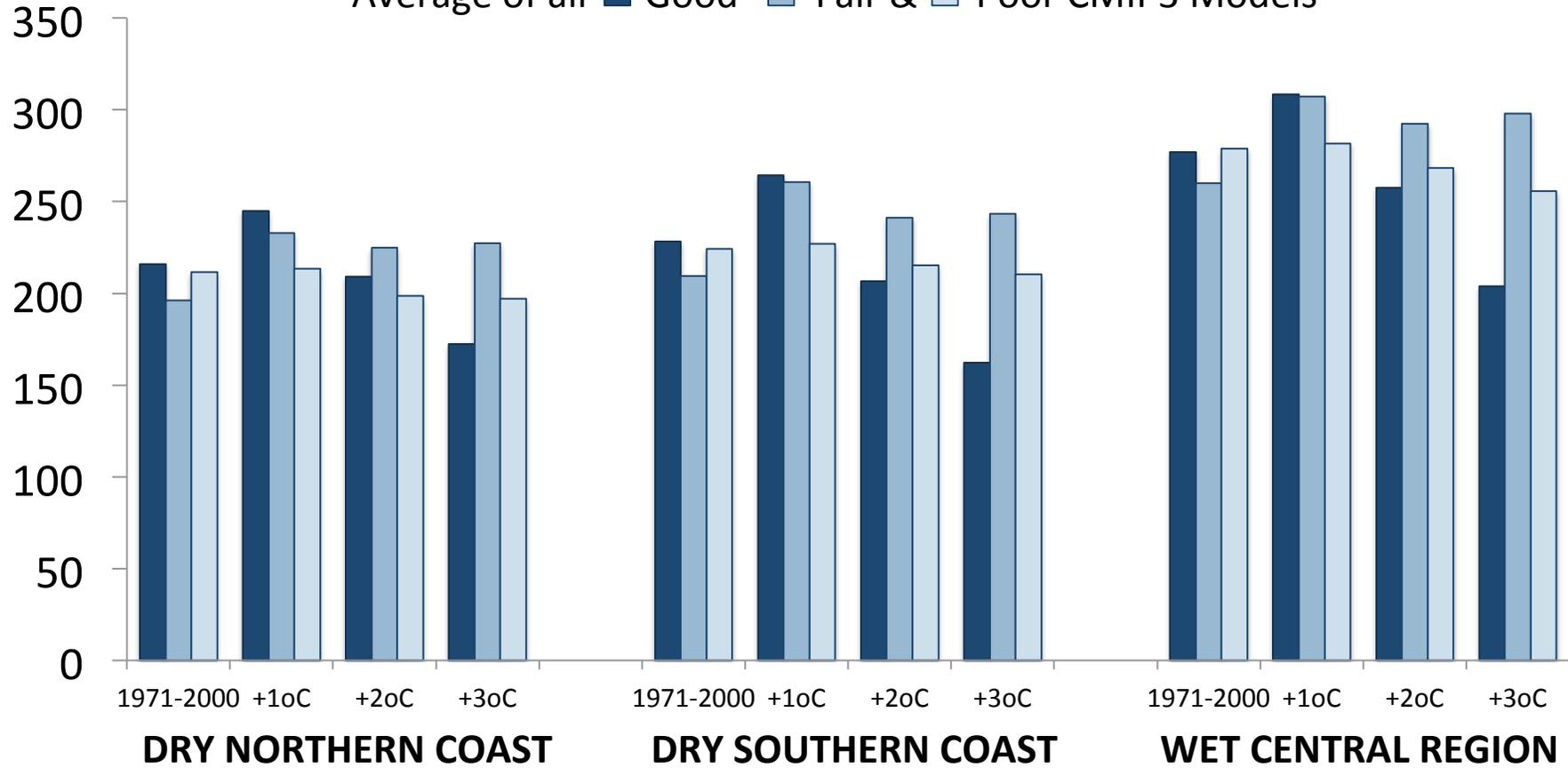
# Precipitation in Wettest 3-Day Period of the Year (mm)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



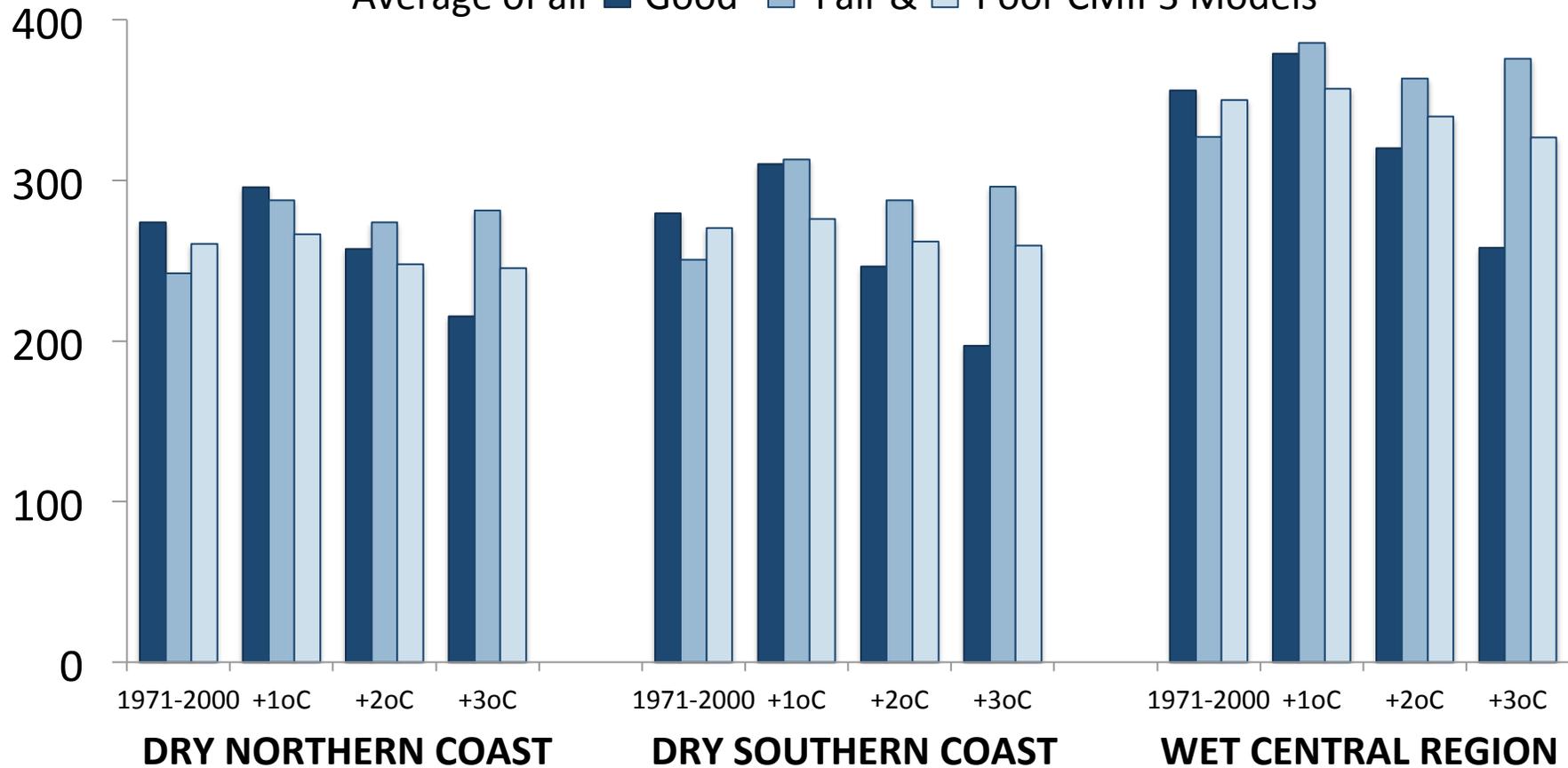
# Precipitation in Wettest 5-Day Period of the Year (mm)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



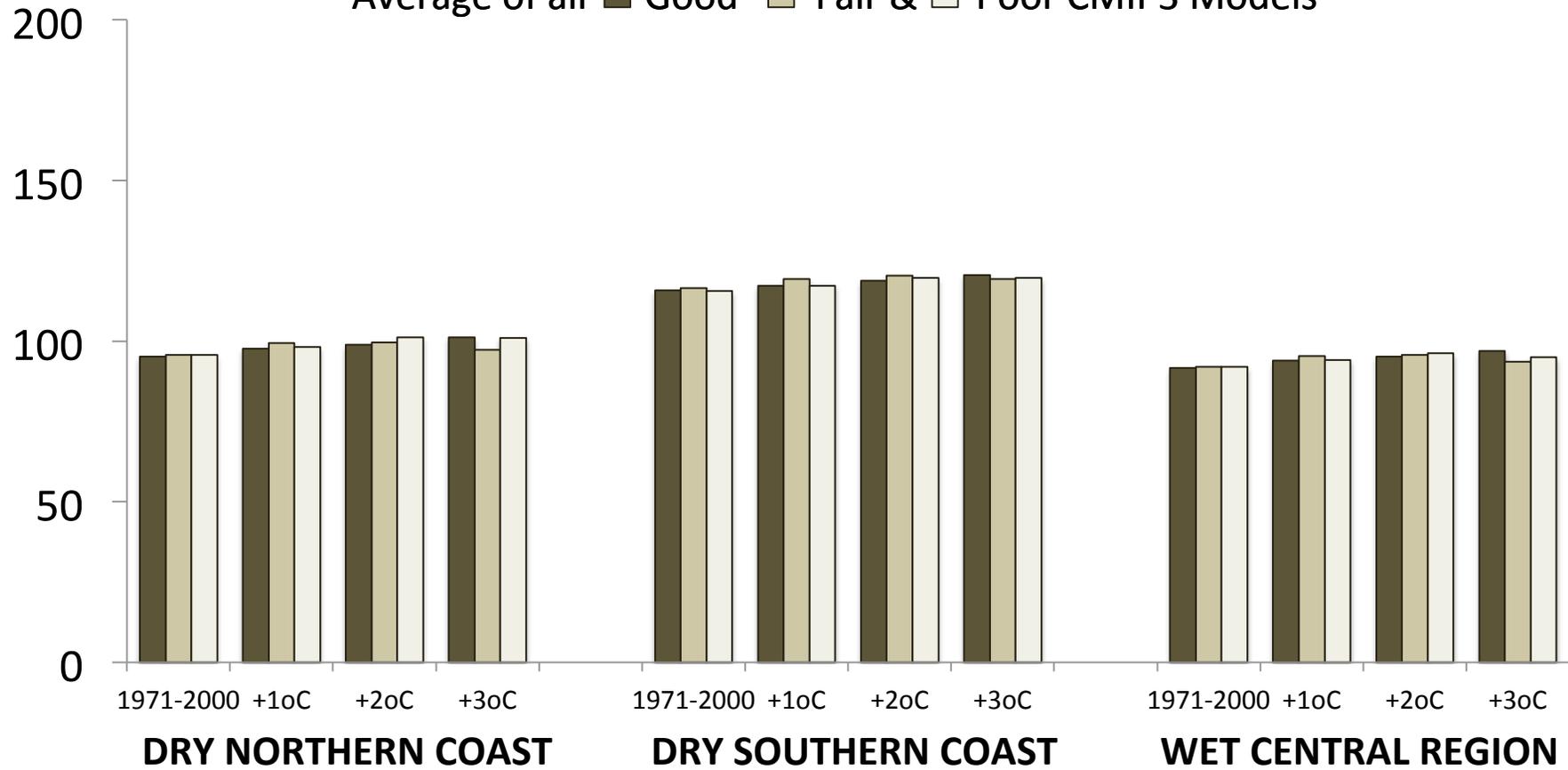
# Precipitation in Wettest 10-Day Period of the Year (mm)

Average of all **Good** Fair & Poor CMIP3 Models



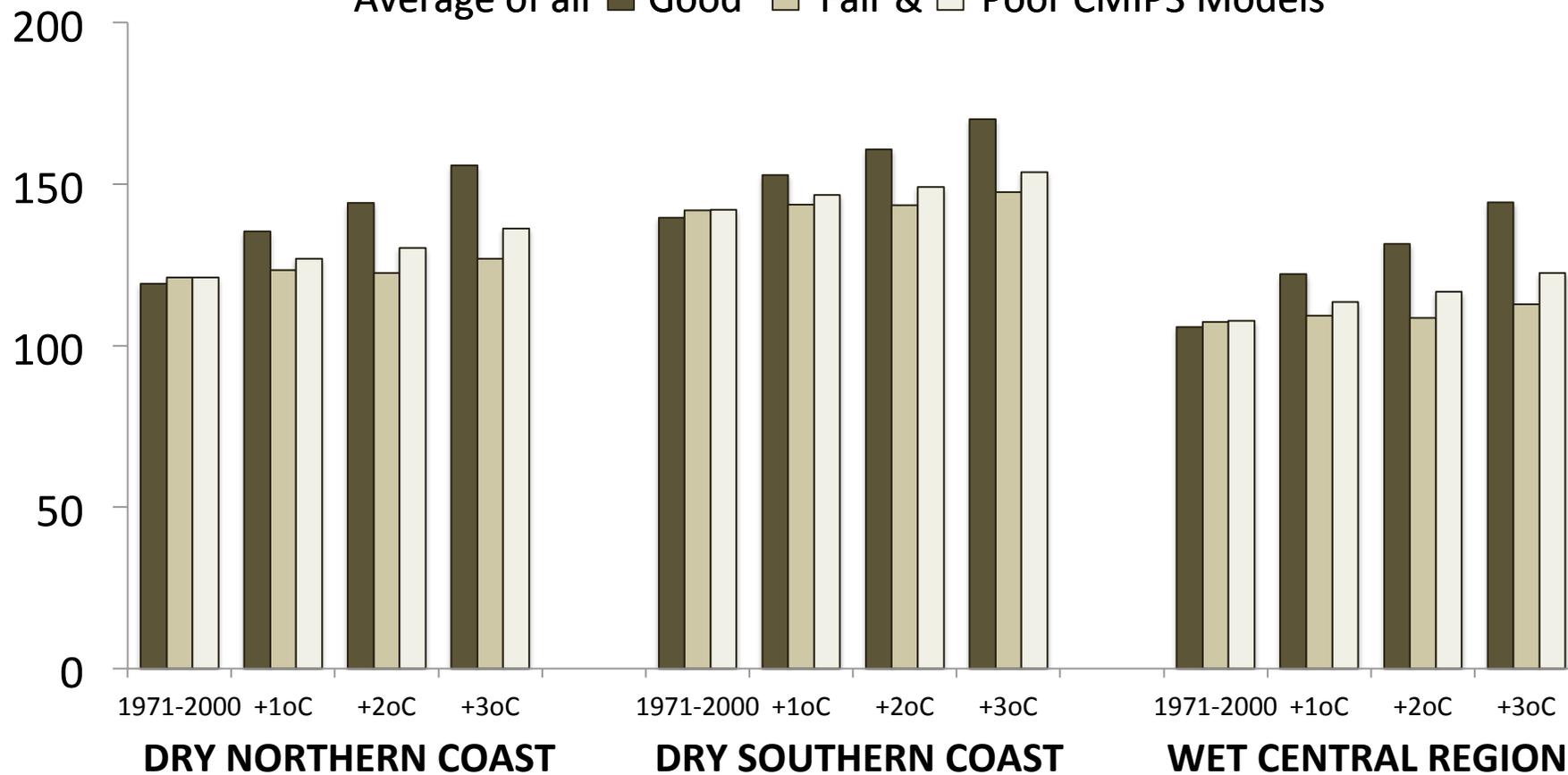
# Number of Dry Days in Dry Season

Average of all  Good  Fair &  Poor CMIP3 Models



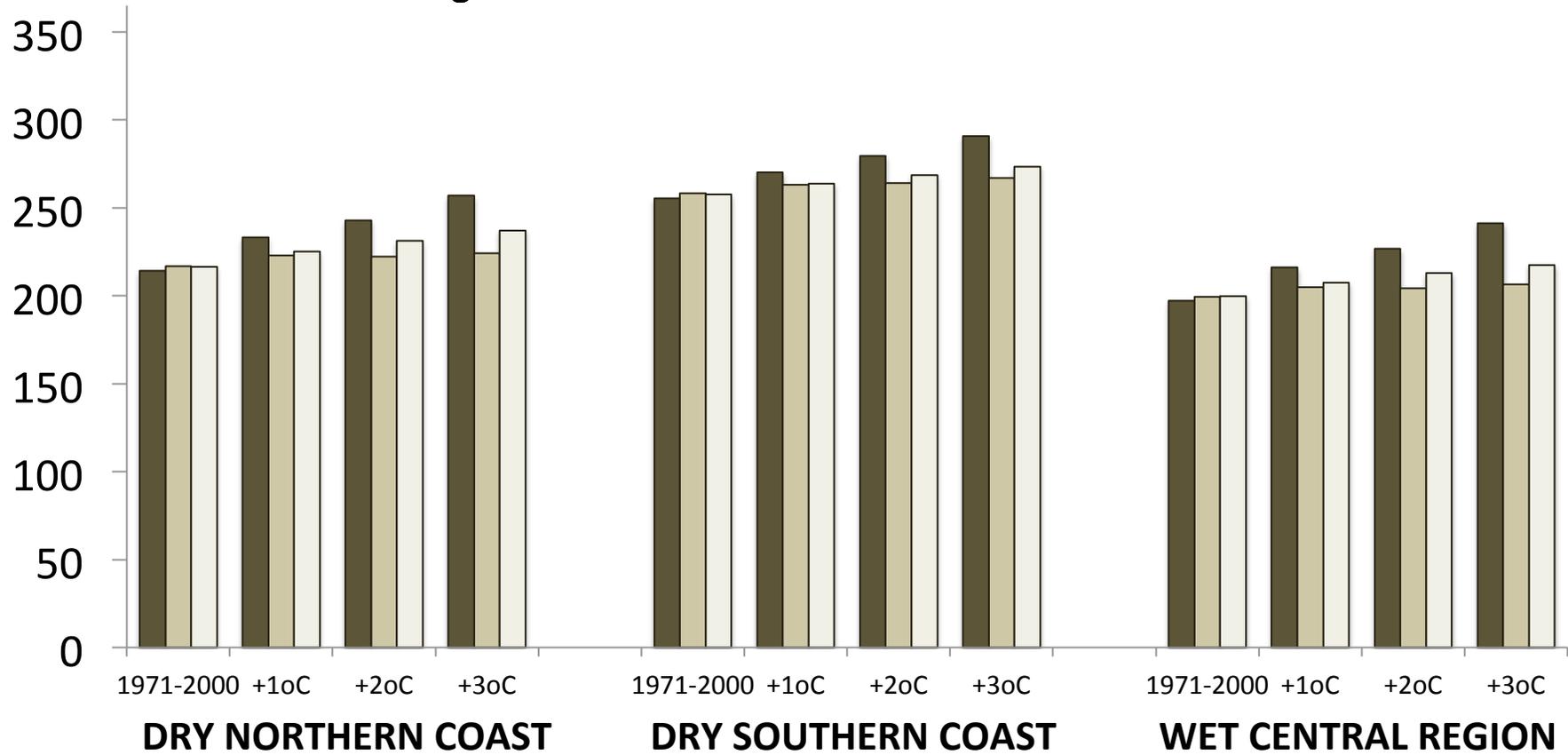
# Number of Dry Days in Wet Season

Average of all **Good** **Fair &** **Poor** CMIP3 Models



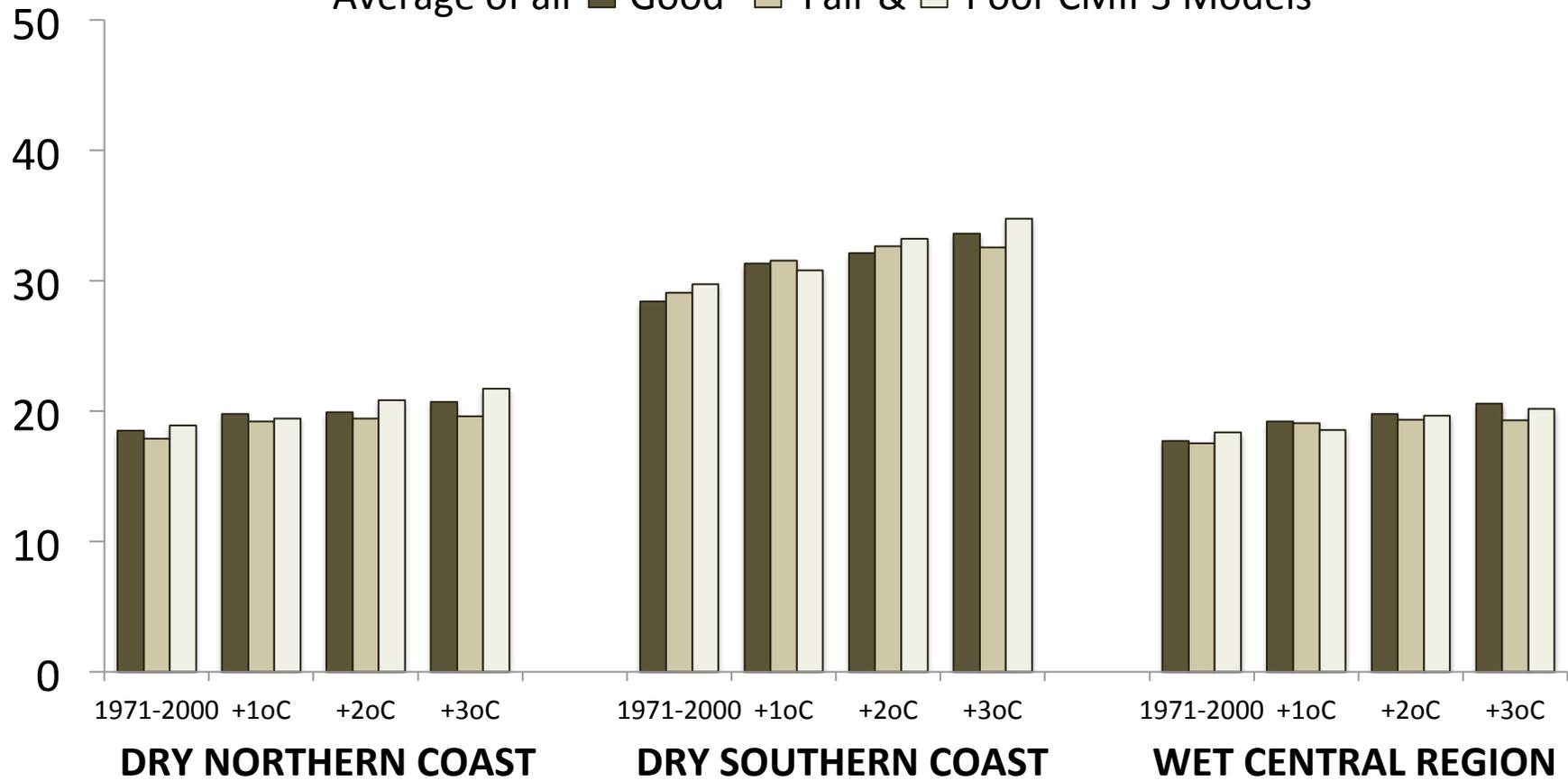
# Number of Dry Days per Year

Average of all **Good** **Fair &** **Poor** CMIP3 Models



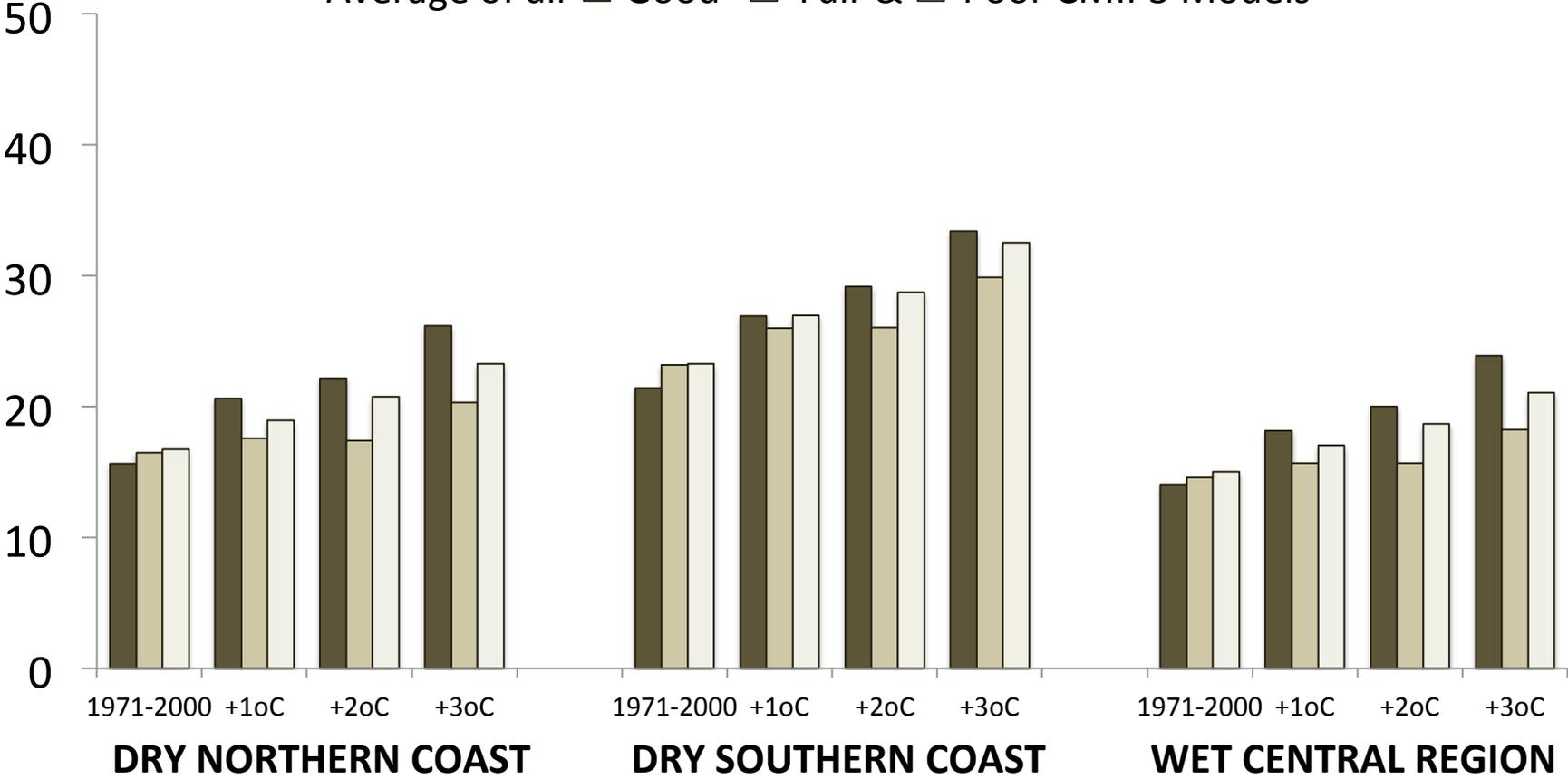
# Longest Dry Period in Dry Season (days)

Average of all  Good  Fair &  Poor CMIP3 Models



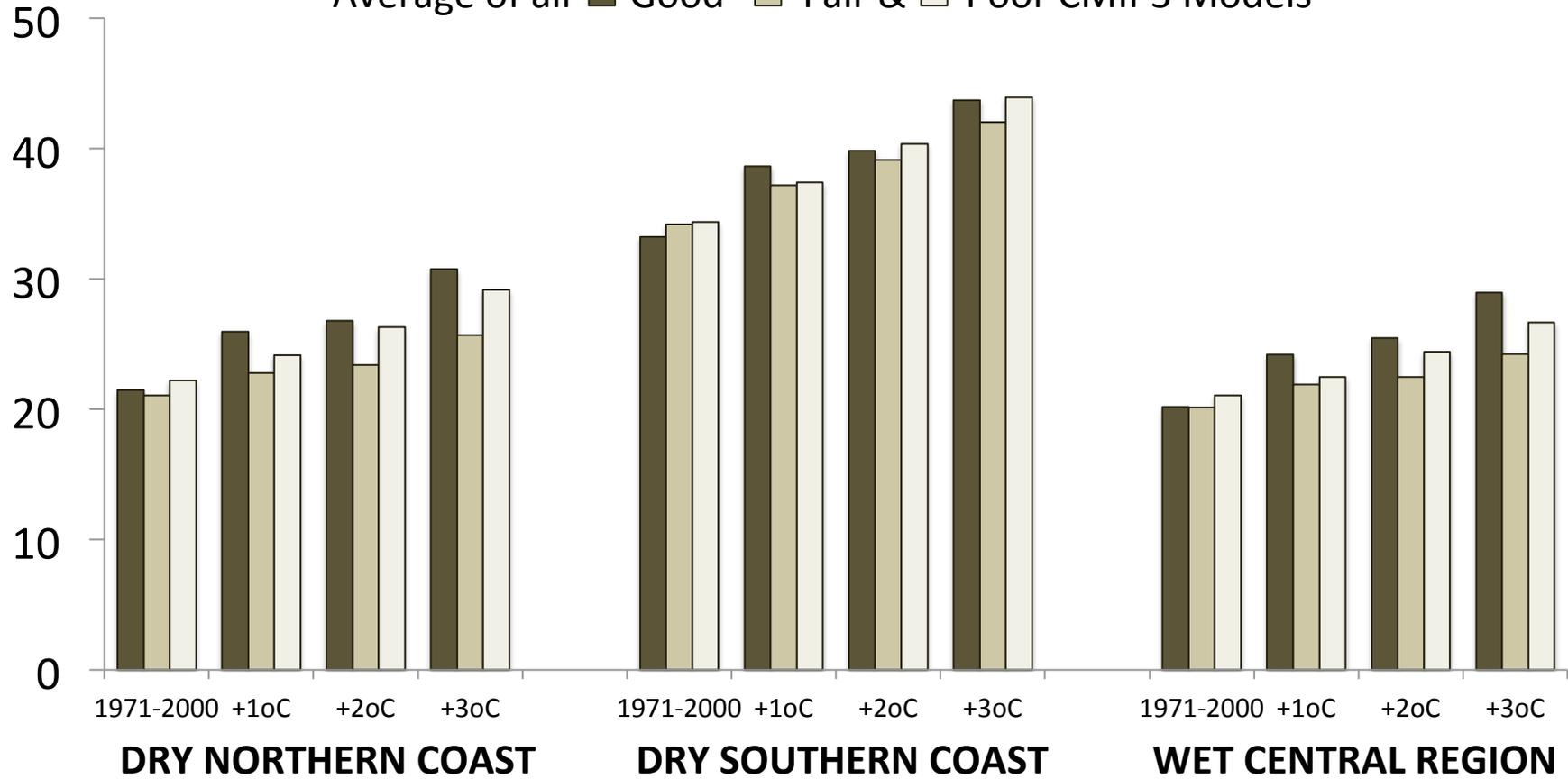
# Longest Dry Period in Wet Season (days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



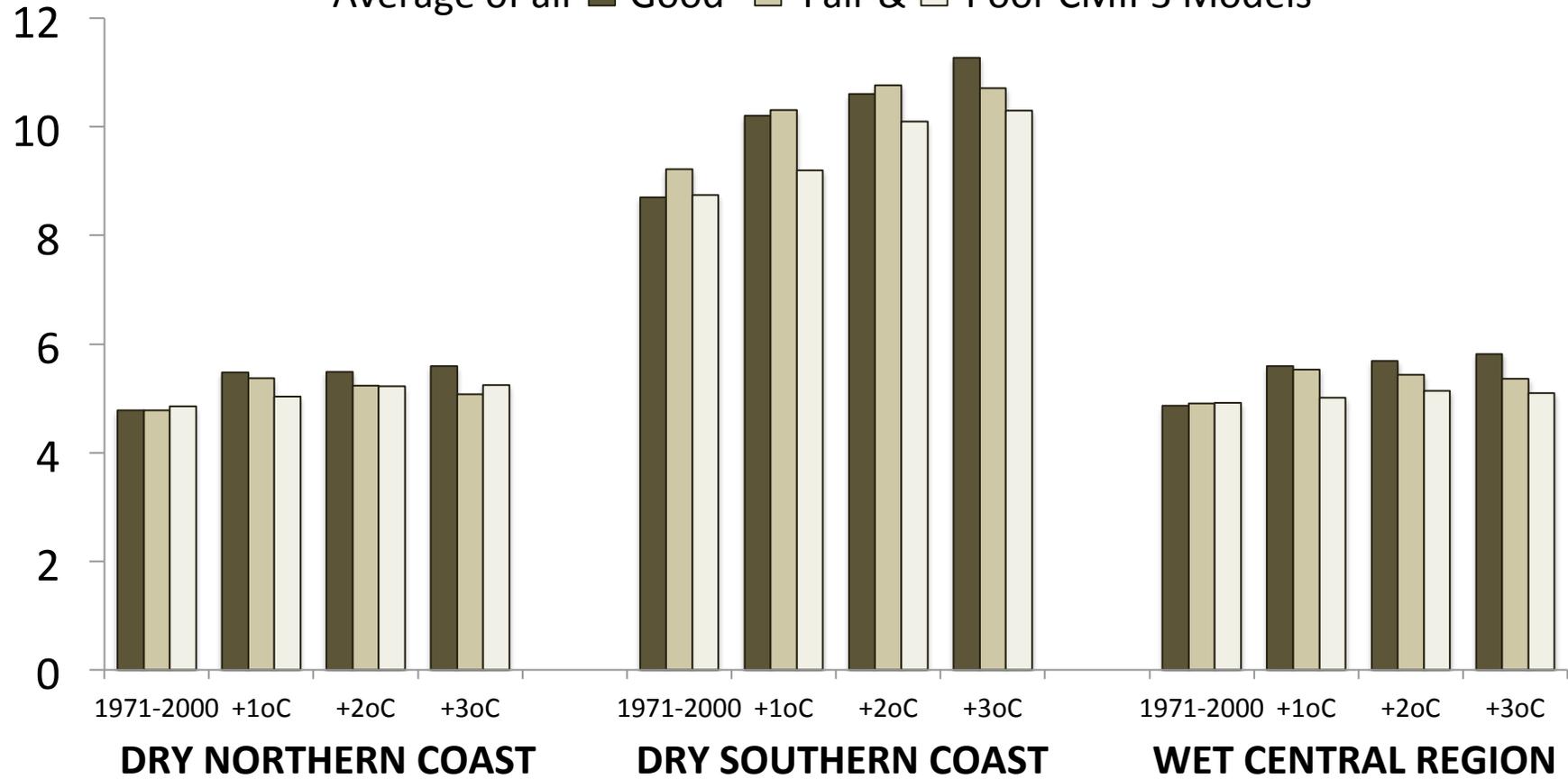
# Longest Dry Period in Year (days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



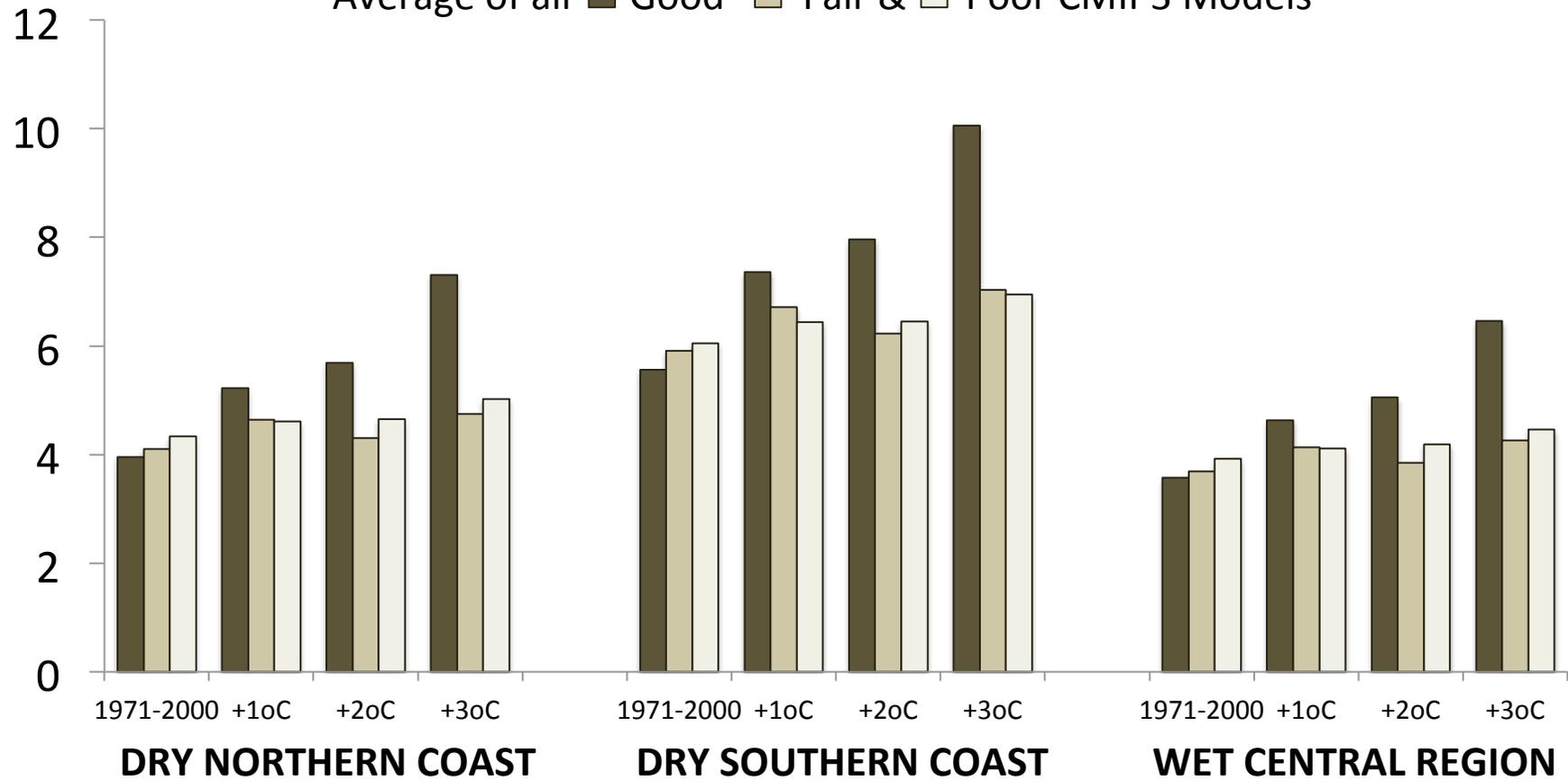
# Average Dry Period in Dry Season (days)

Average of all  Good  Fair &  Poor CMIP3 Models



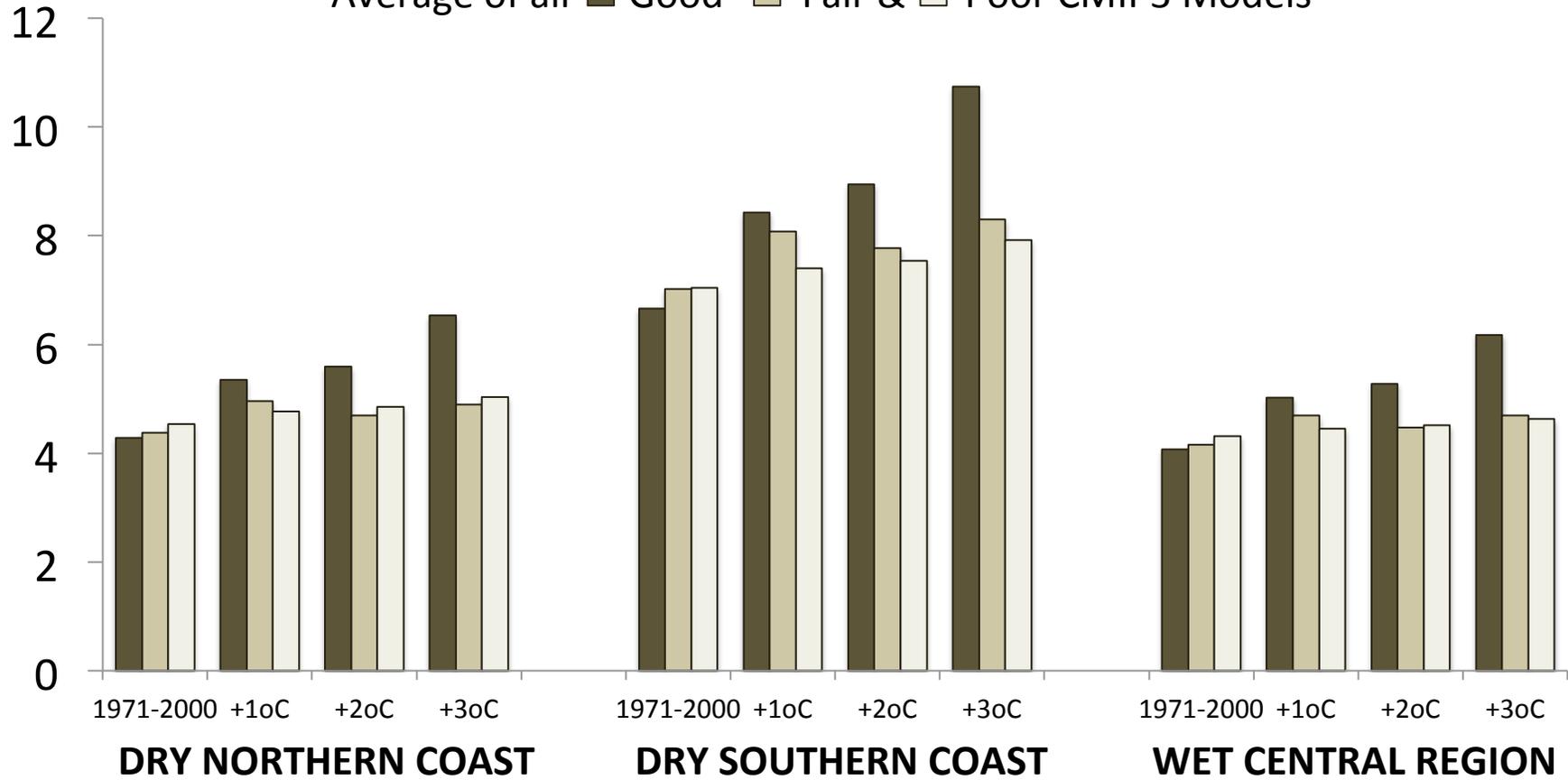
# Average Dry Period in Wet Season (days)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



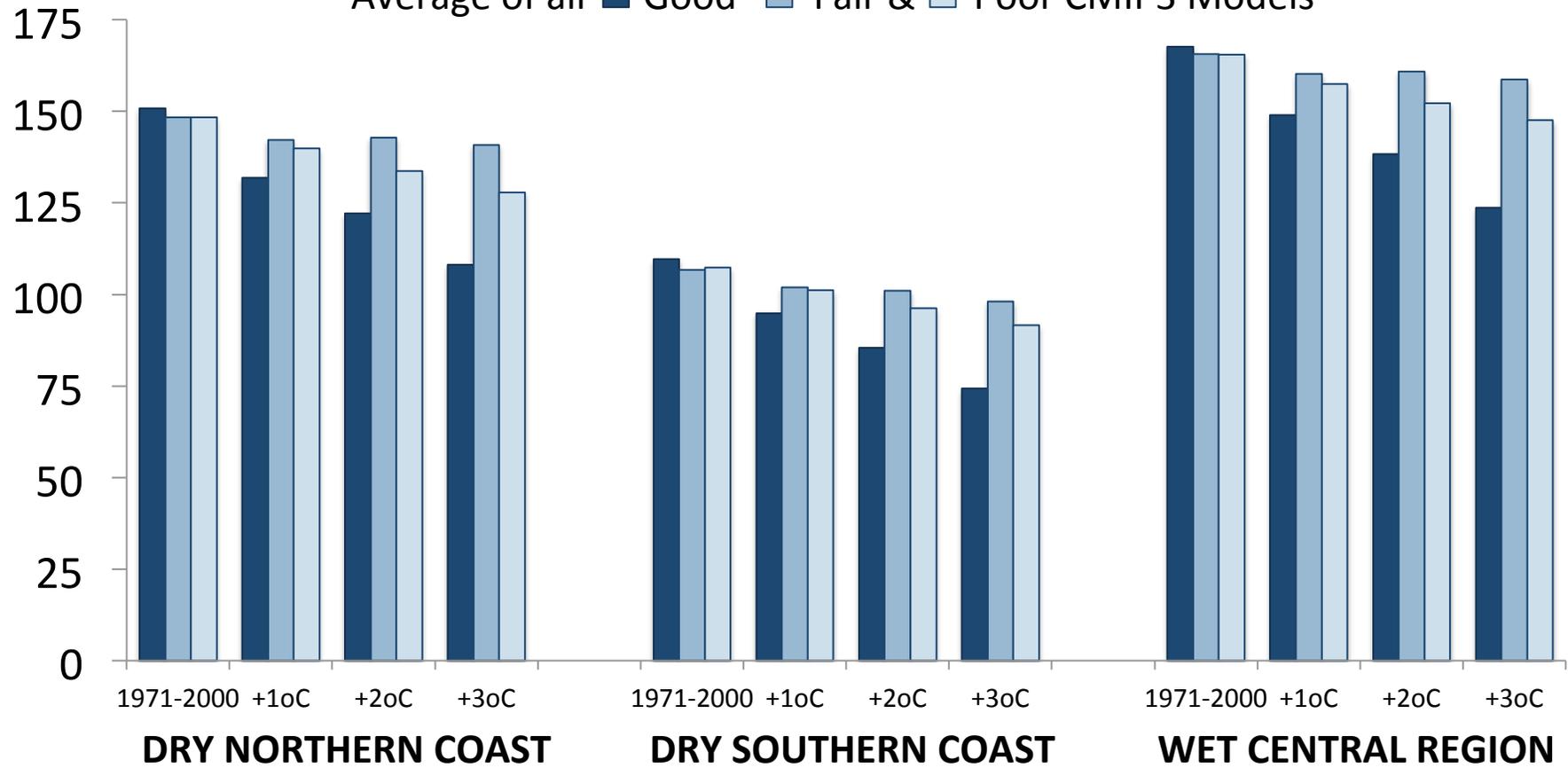
# Average Dry Period in Year (days)

Average of all  Good  Fair &  Poor CMIP3 Models



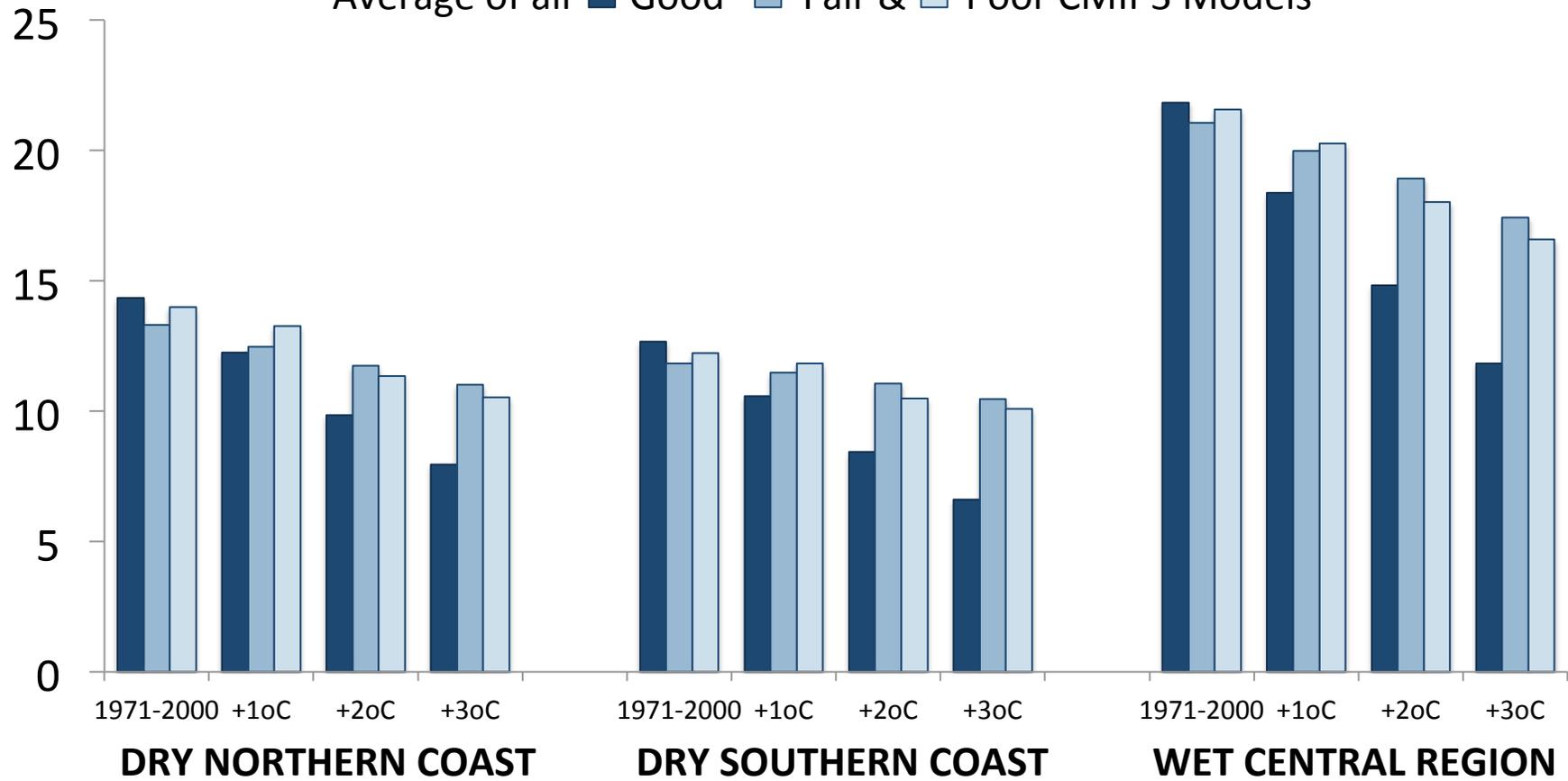
# Number of Wet Days Per Year

Average of all **Good** Fair & Poor CMIP3 Models



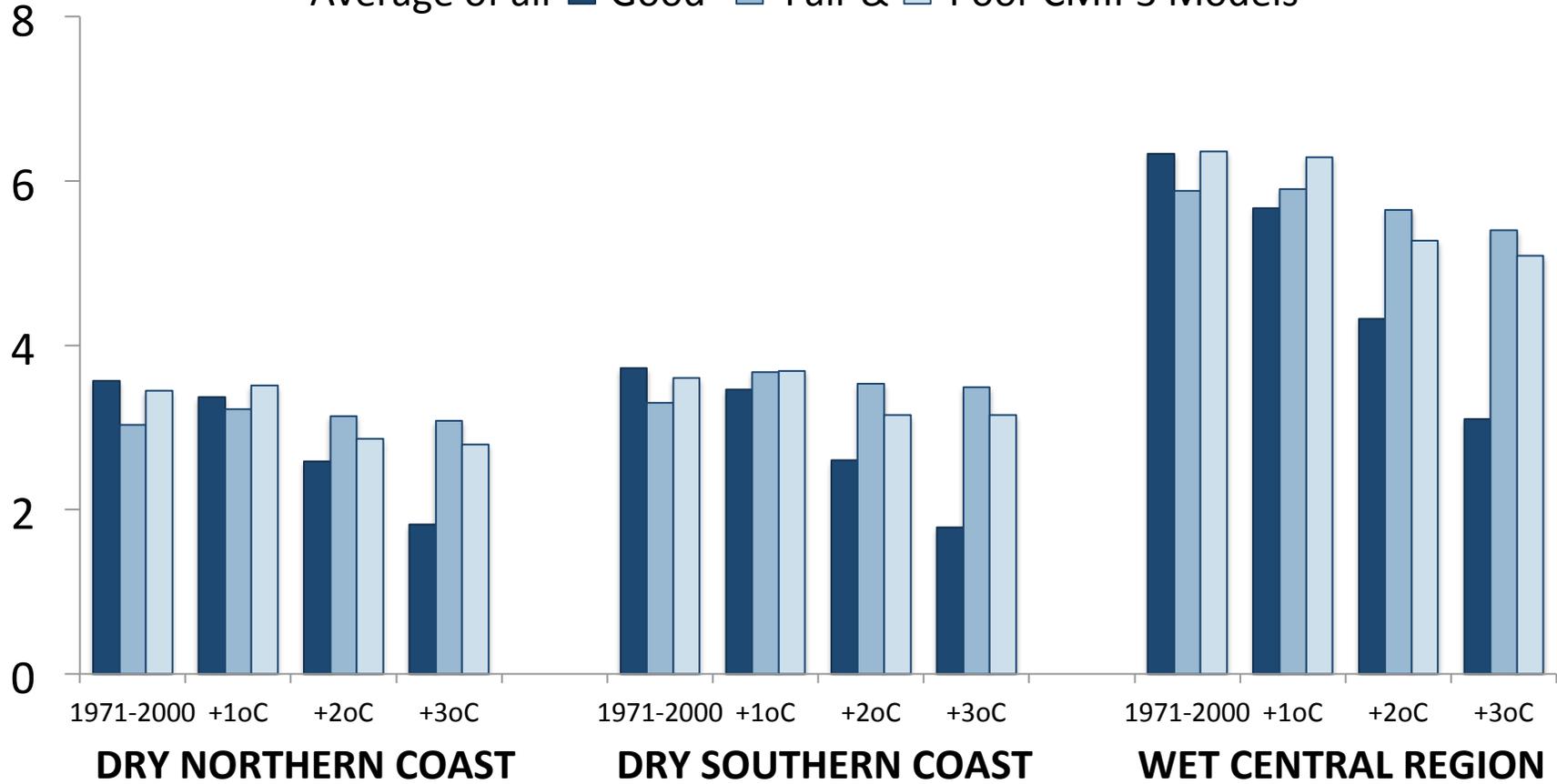
# Precipitation > 1 inch (days per year)

Average of all **Good** Fair & Poor CMIP3 Models



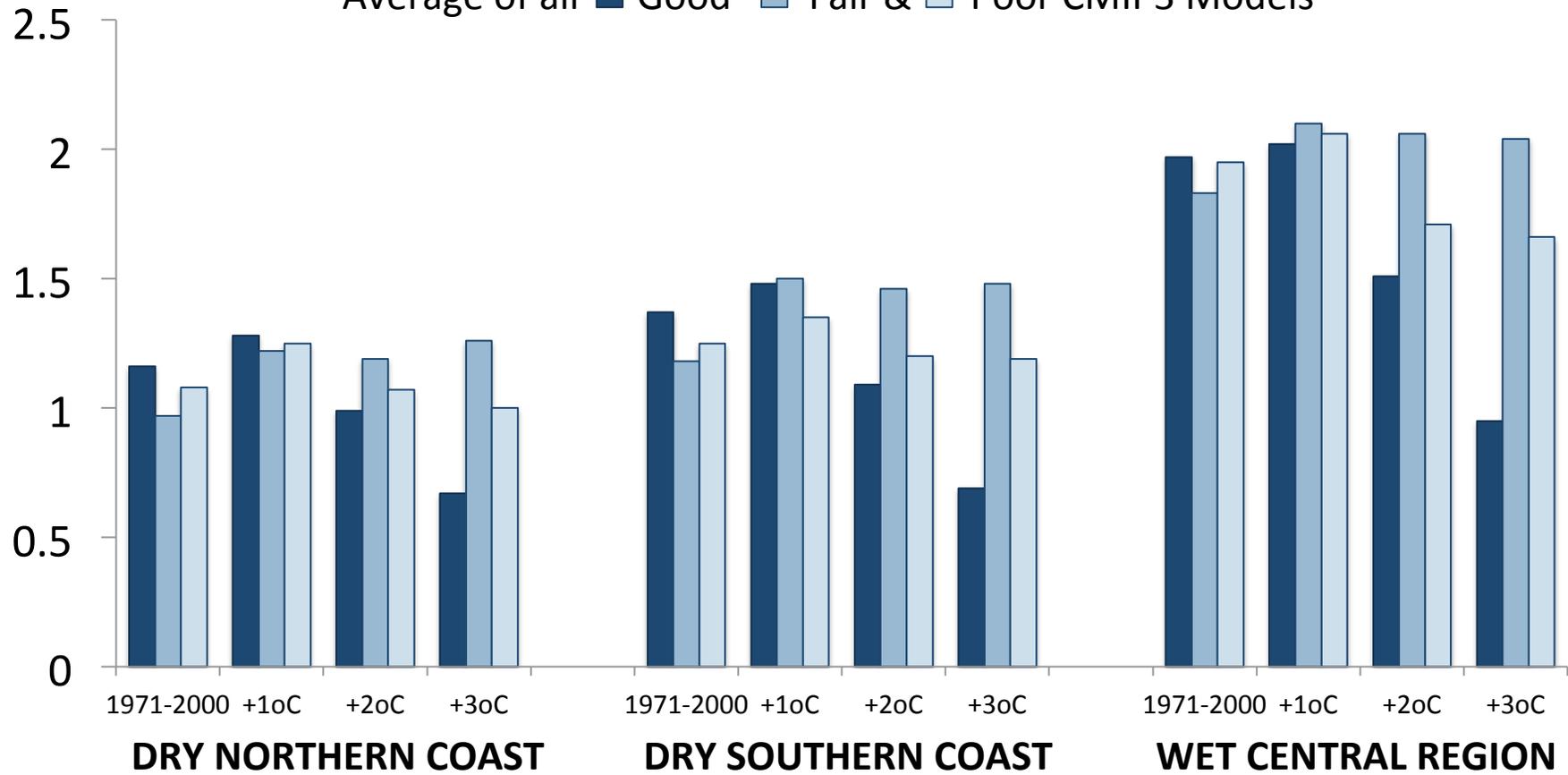
## Precipitation > 2 inches (days per year)

Average of all **Good** Fair & Poor CMIP3 Models



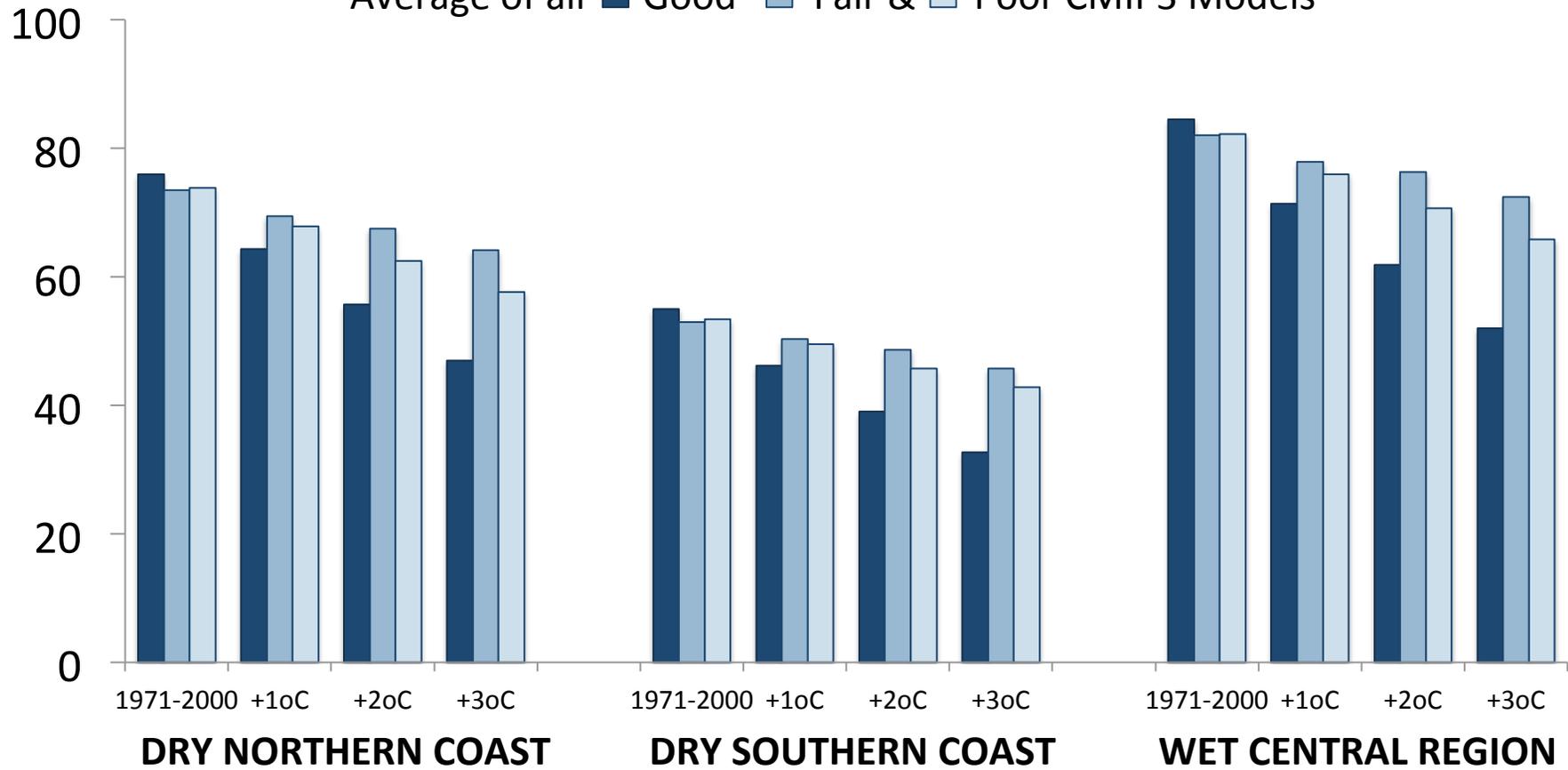
# Precipitation > 3 inches (days per year)

Average of all ■ Good ■ Fair & ■ Poor CMIP3 Models



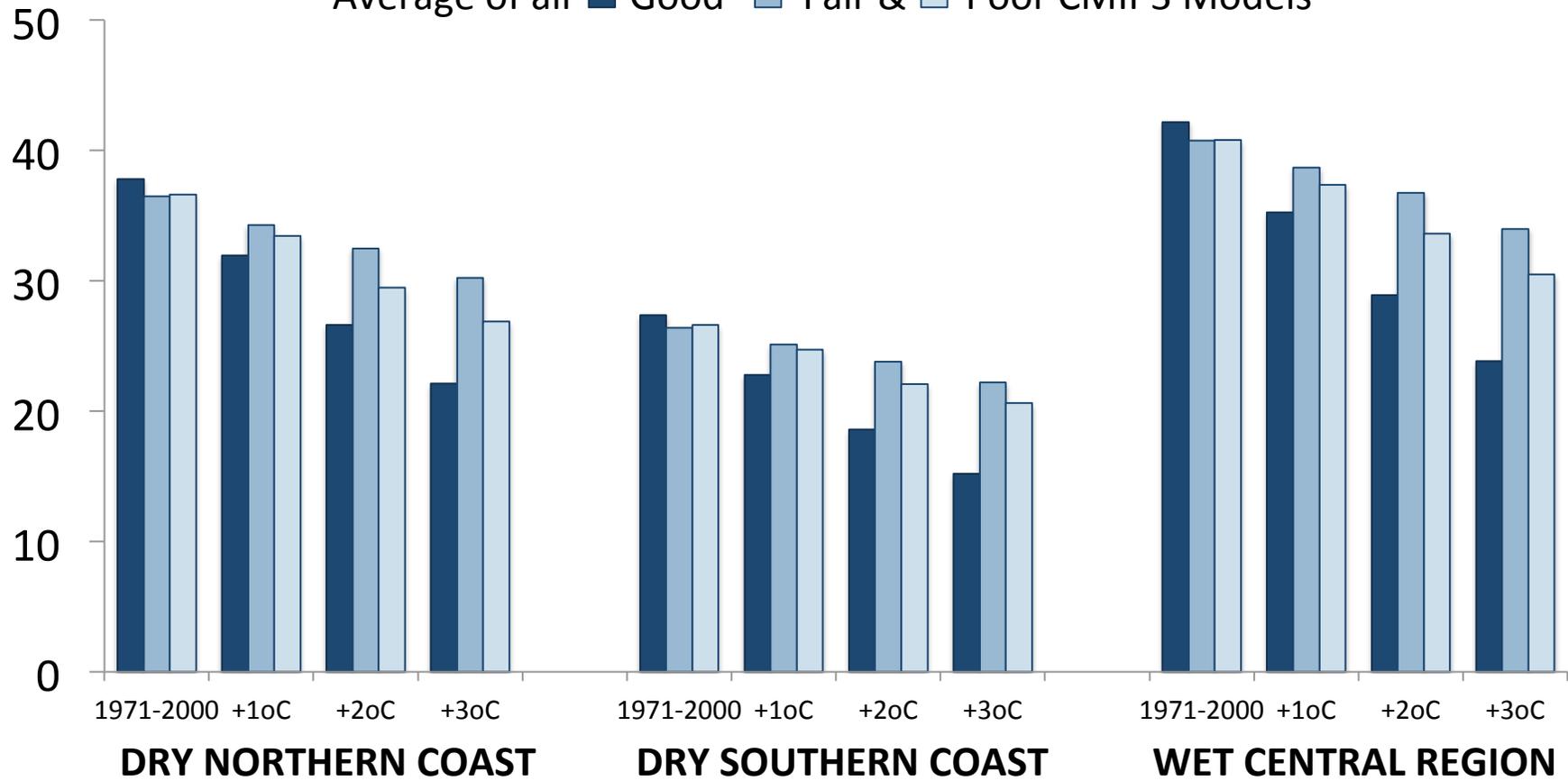
# Precipitation > Historical Median (days per year)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



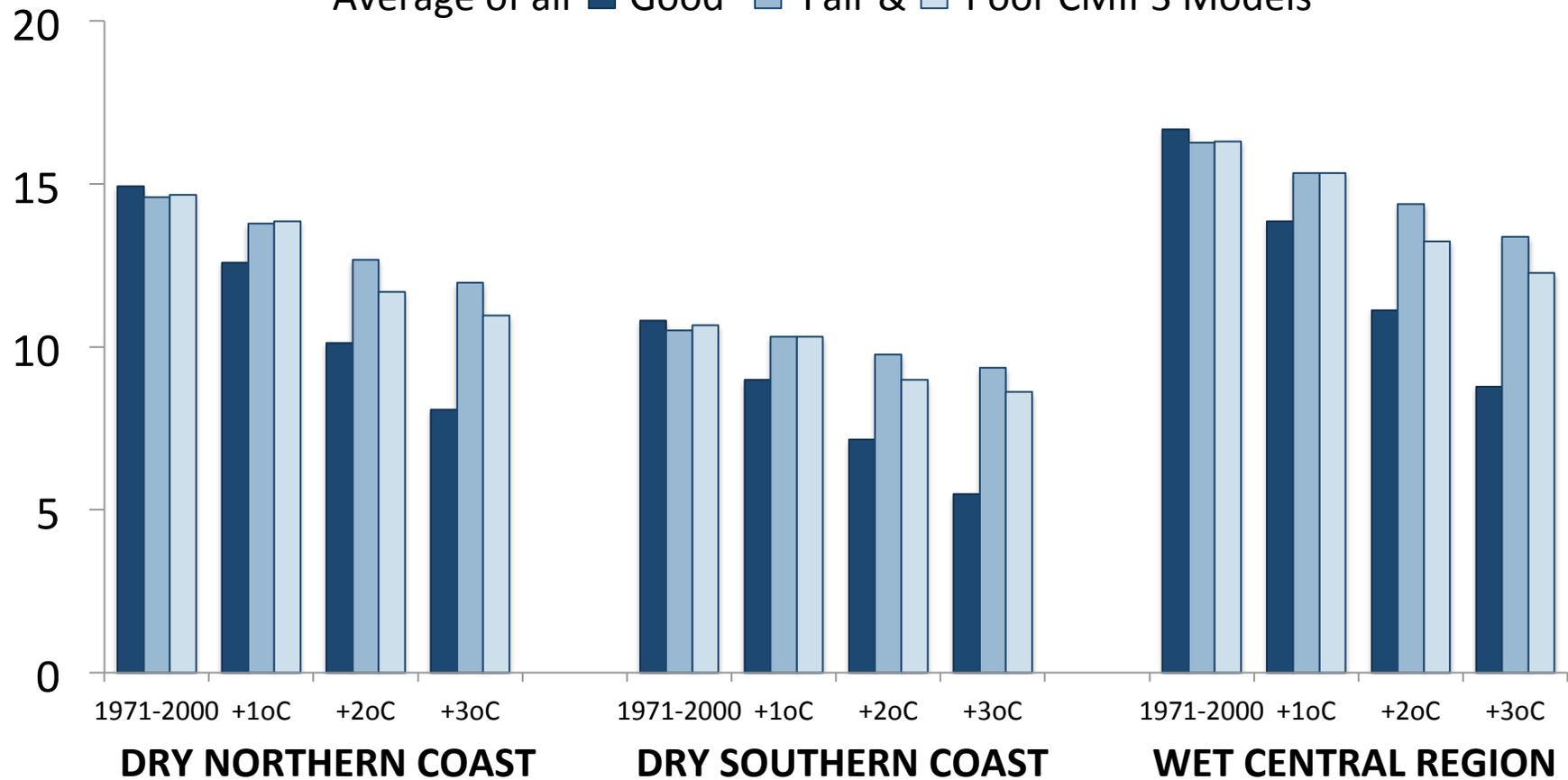
# Precipitation > Historical 75th Percentile (days per year)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



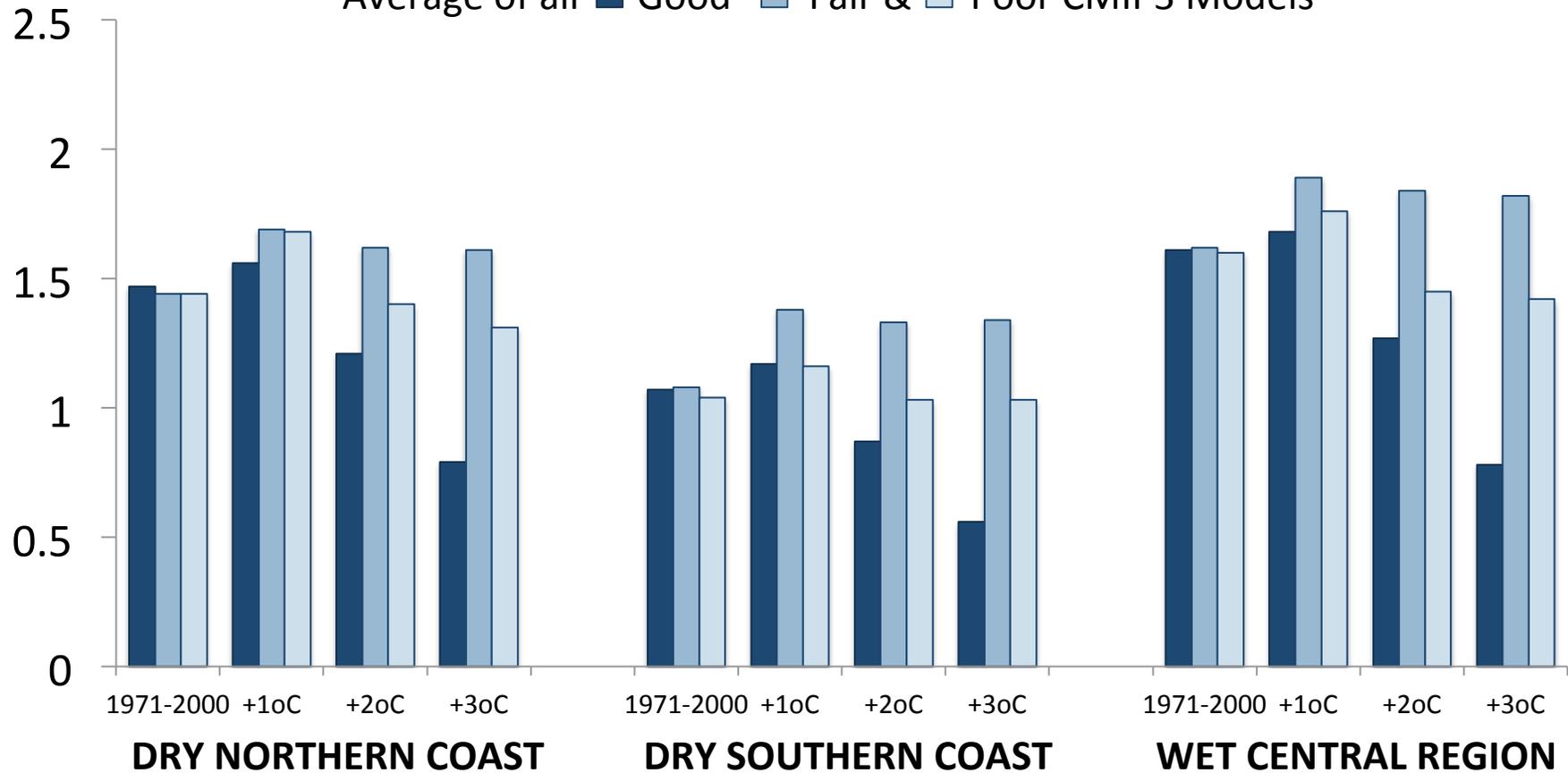
# Precipitation > Historical 90th Percentile (days per year)

Average of all **Good** **Fair &** **Poor** CMIP3 Models



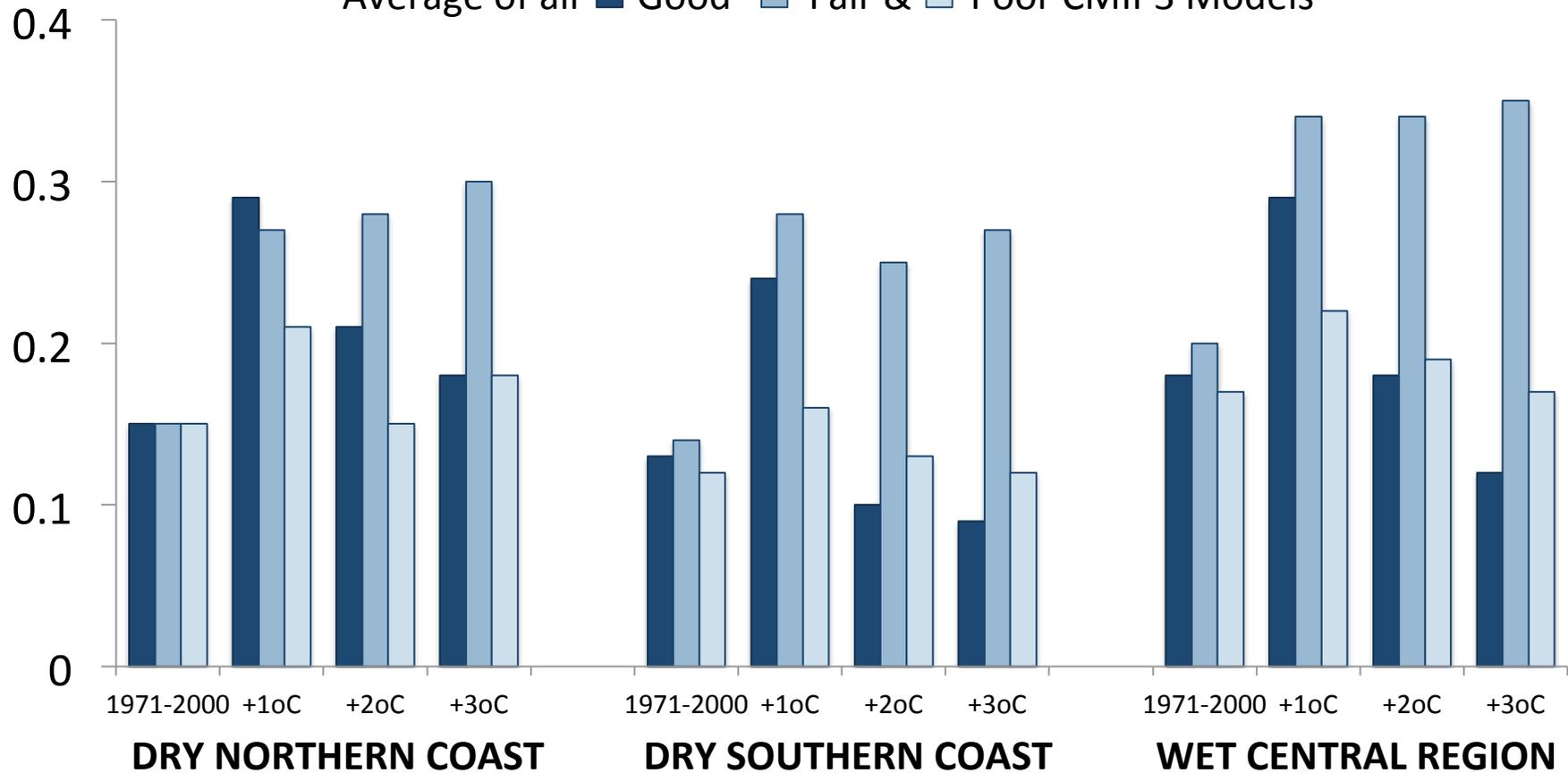
# Precipitation > Historical 99th Percentile (days per year)

Average of all **Good** **Fair &** **Poor** CMIP3 Models

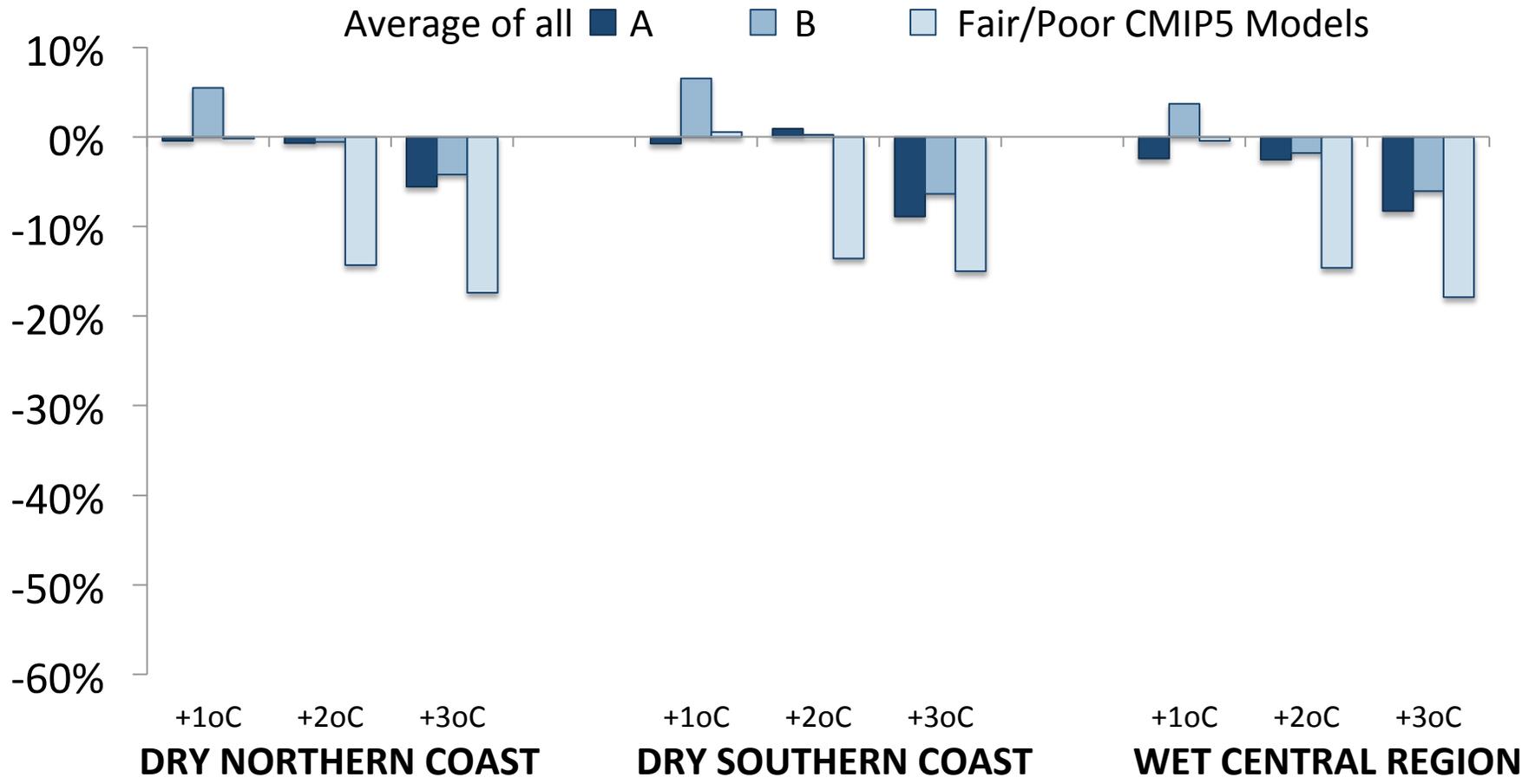


# Precipitation > Historical 99.9th Percentile (days per year)

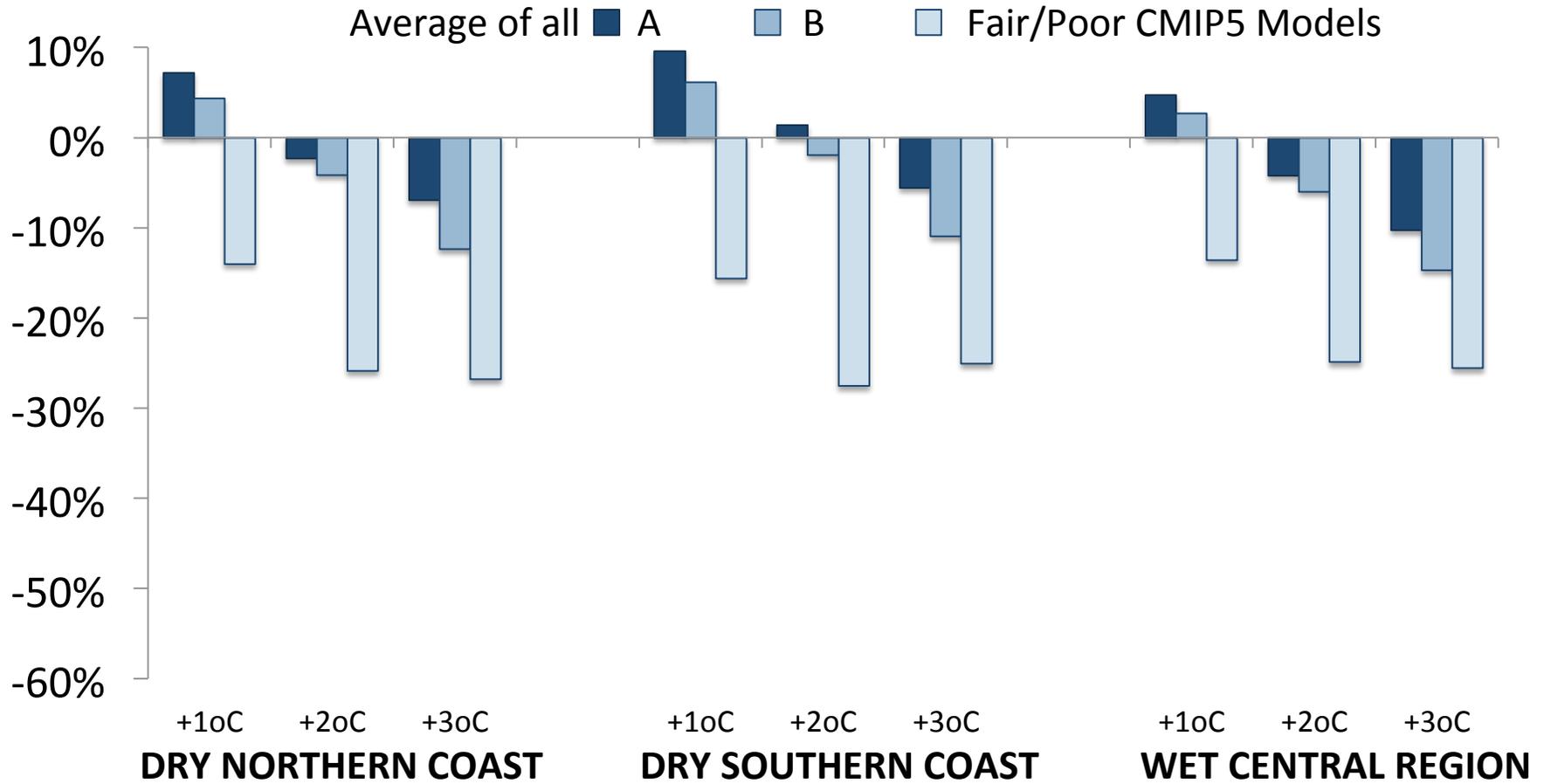
Average of all **Good** **Fair &** **Poor** CMIP3 Models



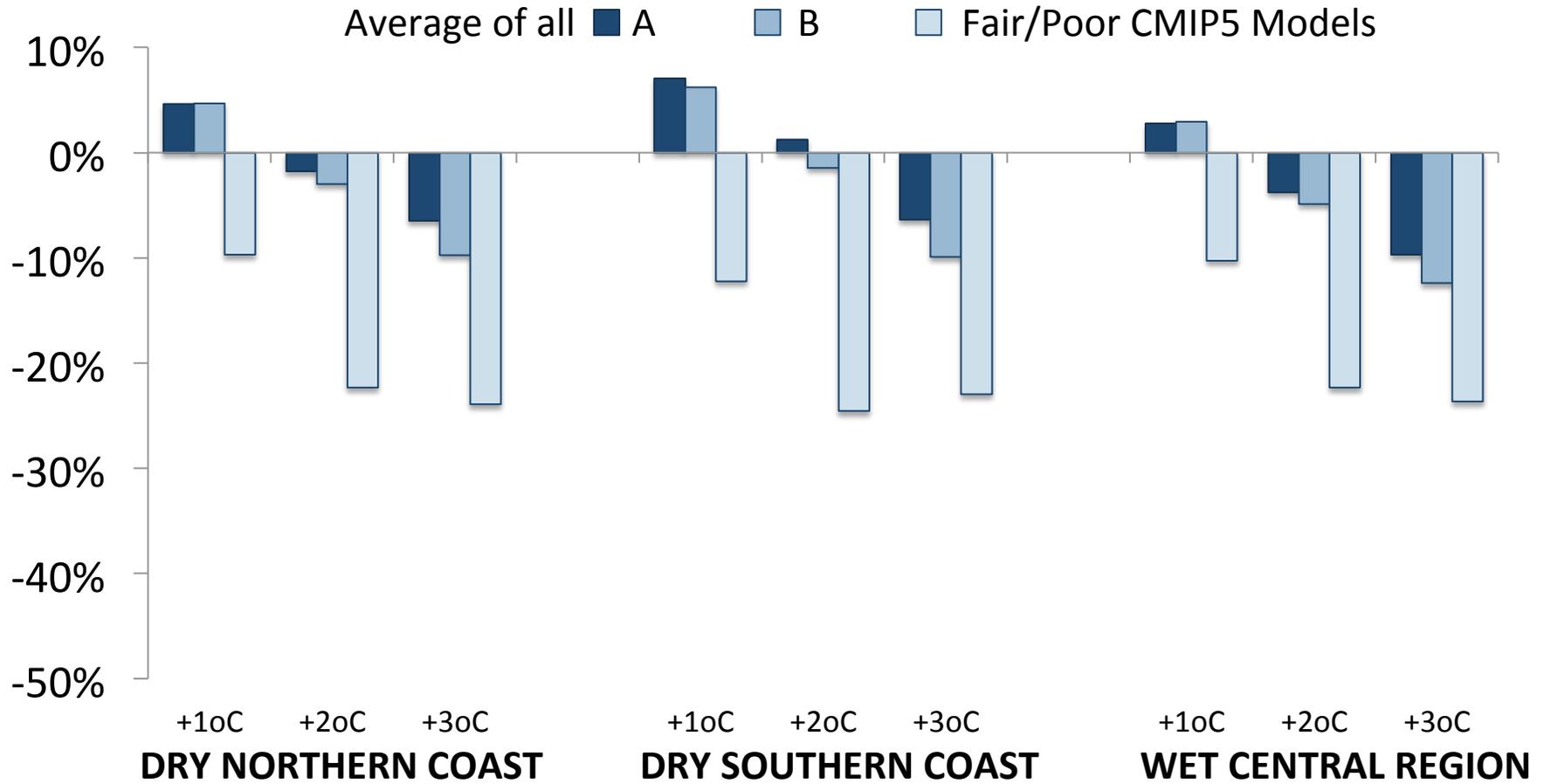
# Precipitation in Dry Season (percentage change)



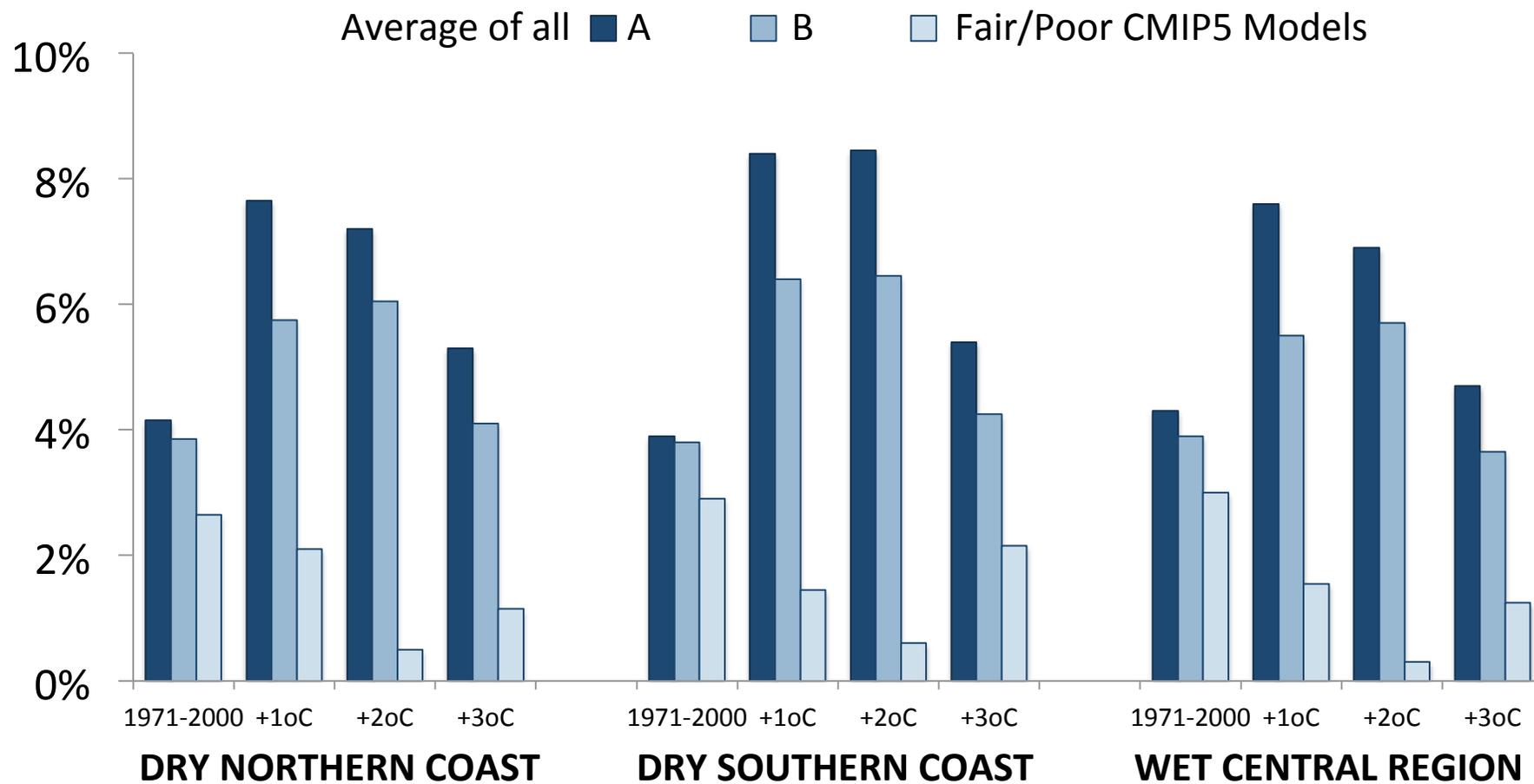
# Precipitation in Wet Season (percentage change)



# Annual Precipitation (percentage change)

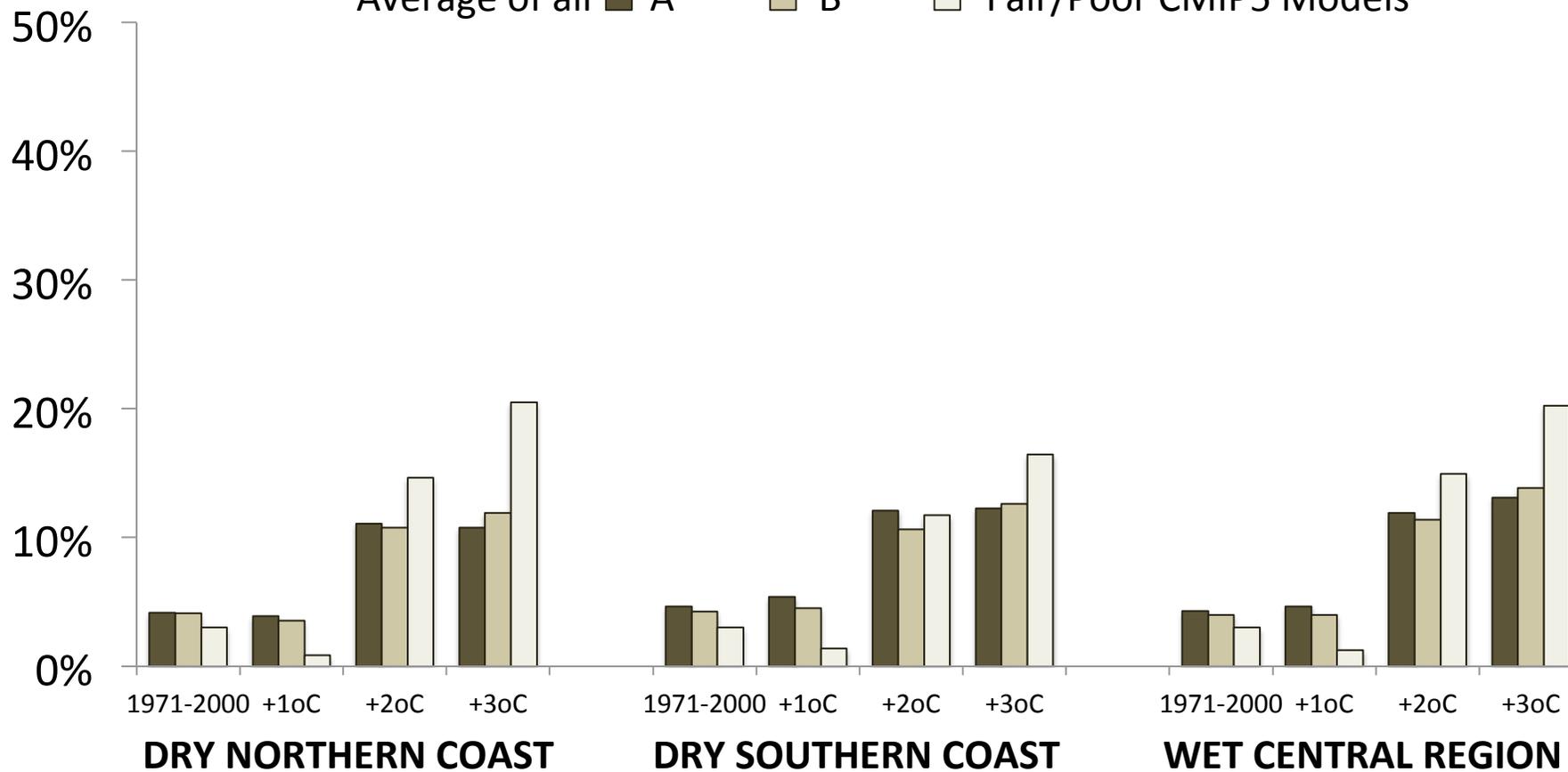


# Wet Years (percent)



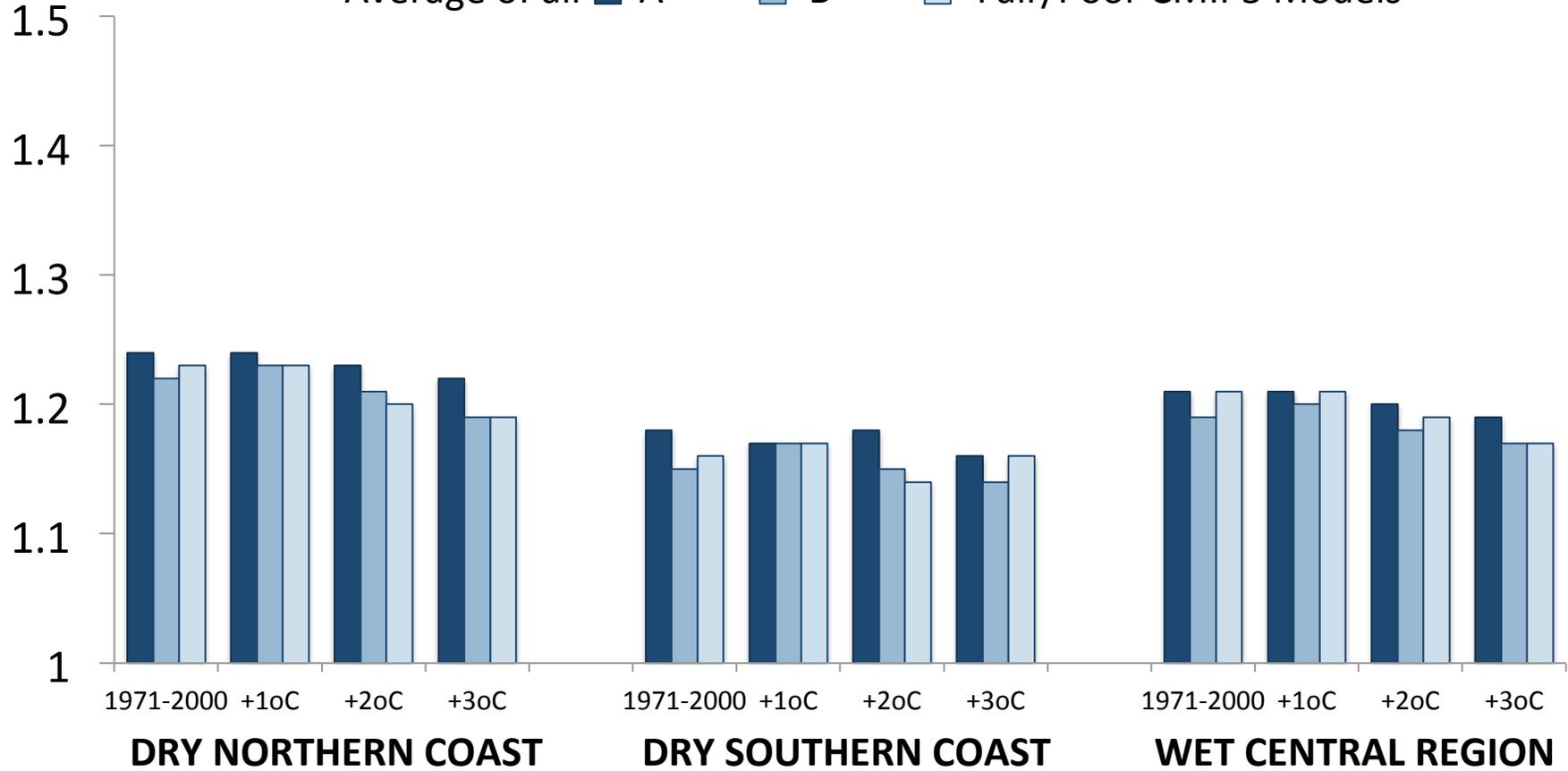
# Dry Years (percent)

Average of all  A  B  Fair/Poor CMIP5 Models



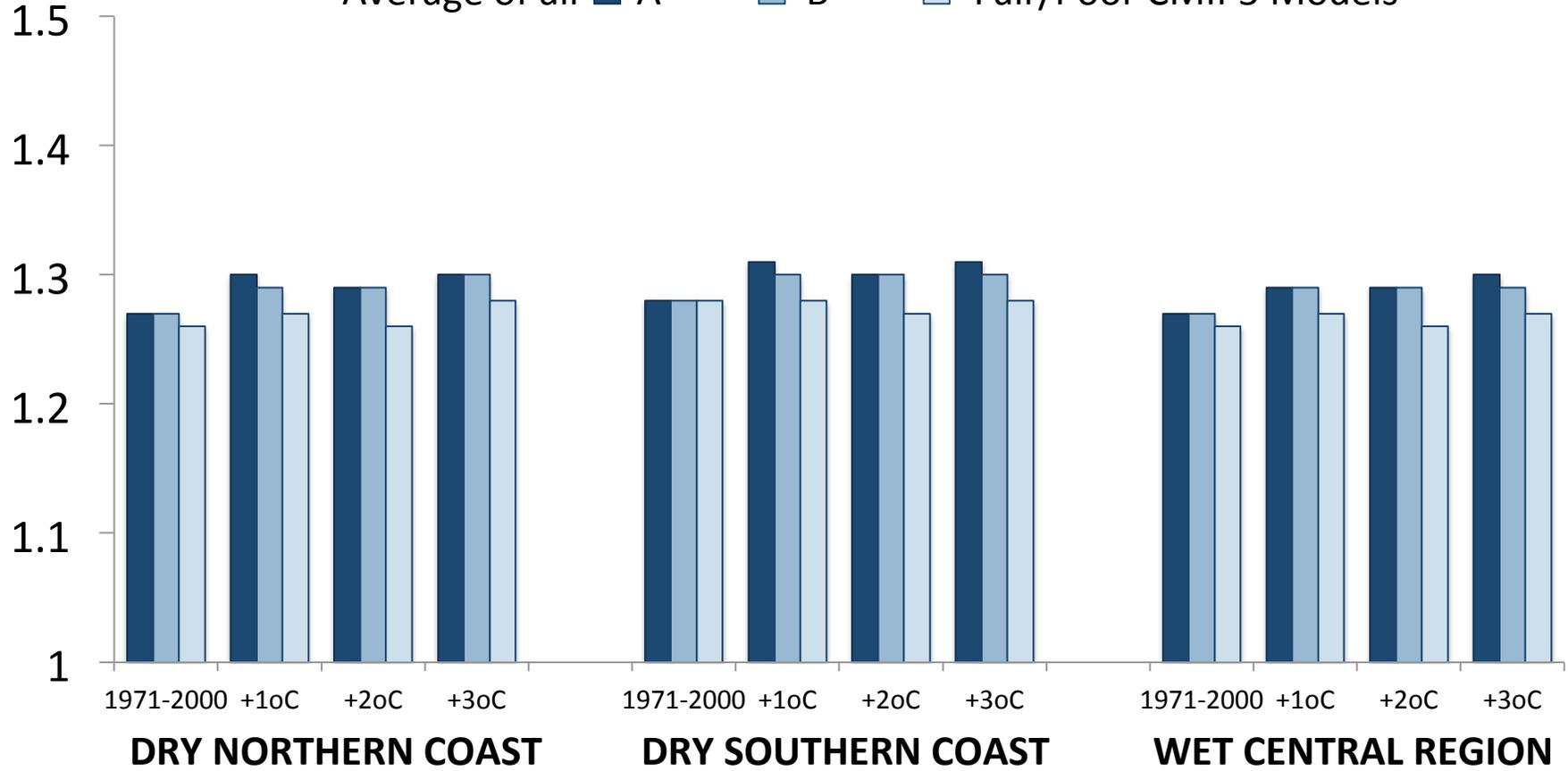
# Standard Deviation of Precipitation in Dry Season (mm)

Average of all ■ A ■ B ■ Fair/Poor CMIP5 Models



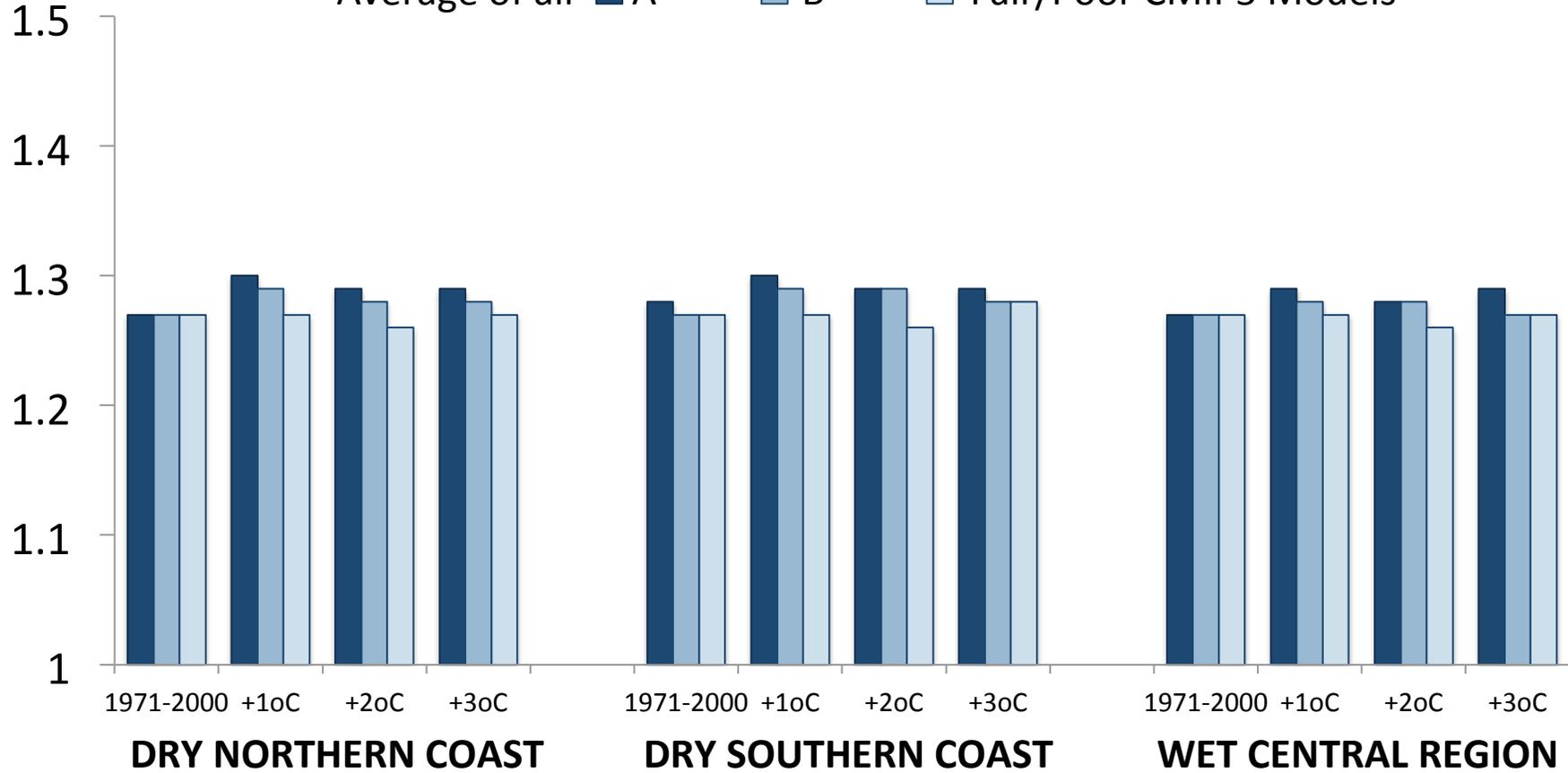
# Standard Deviation of Precipitation in Wet Season (mm)

Average of all **A**    **B**    Fair/Poor CMIP5 Models

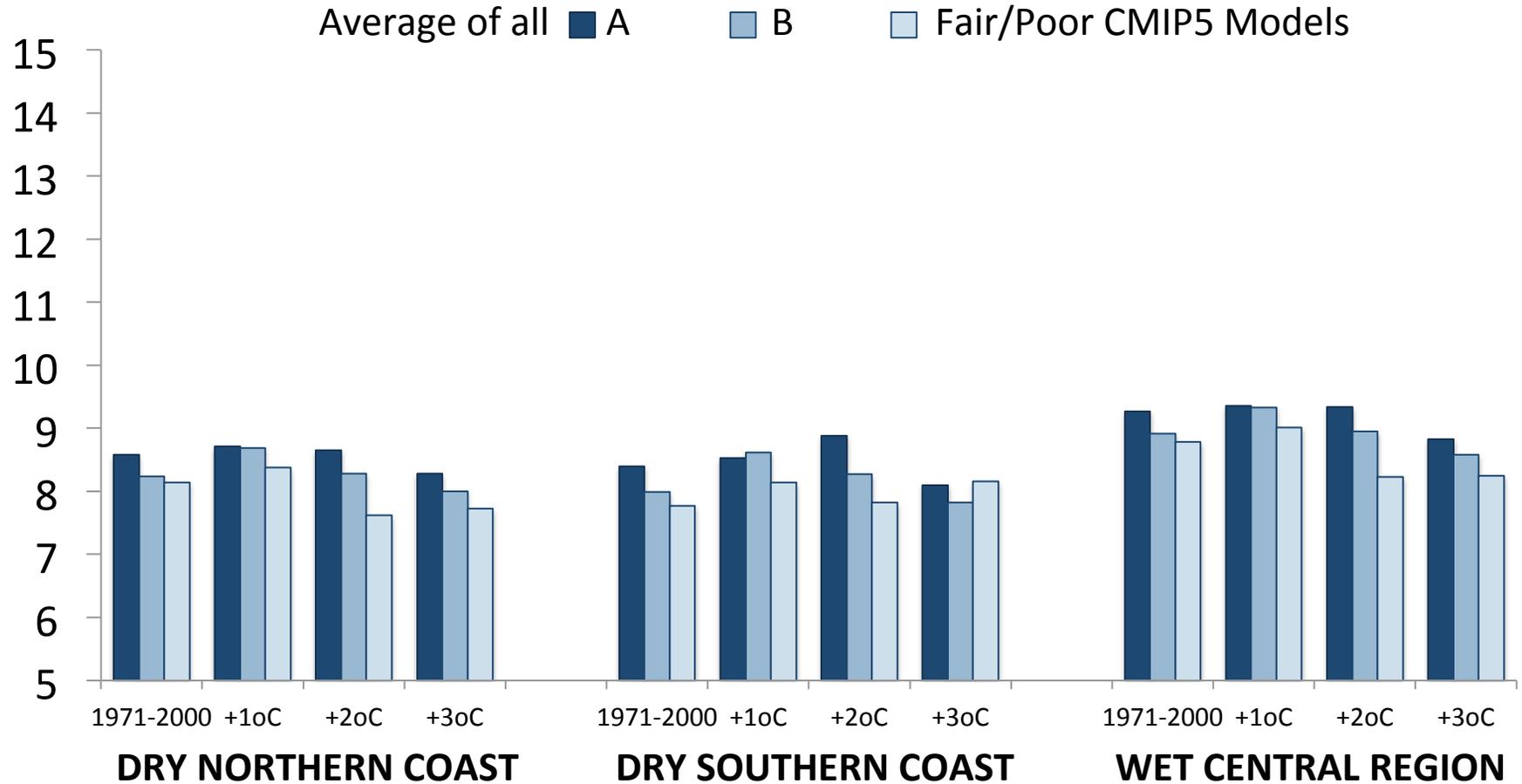


# Standard Deviation of Precipitation Annual (mm)

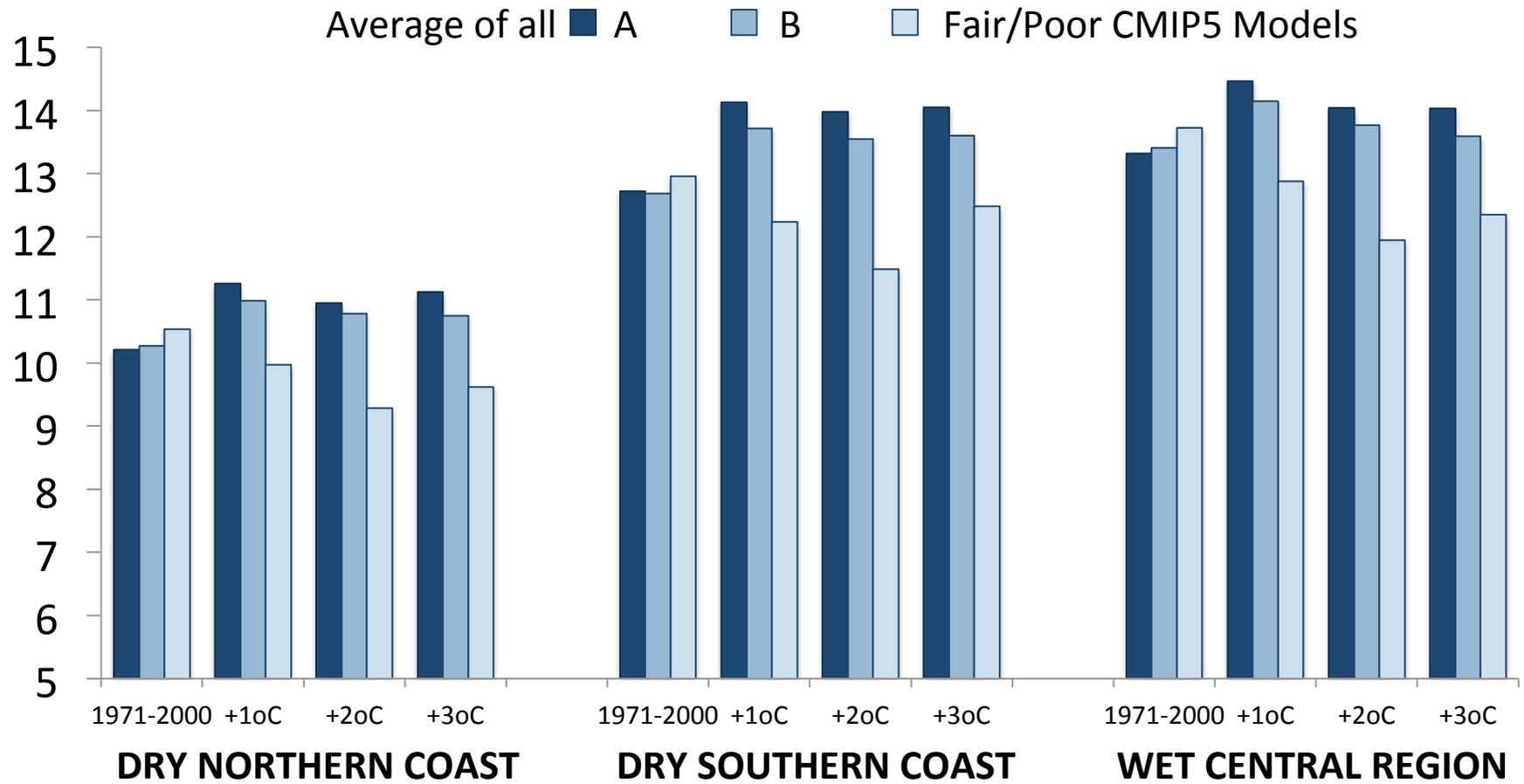
Average of all **A**    **B**    Fair/Poor CMIP5 Models



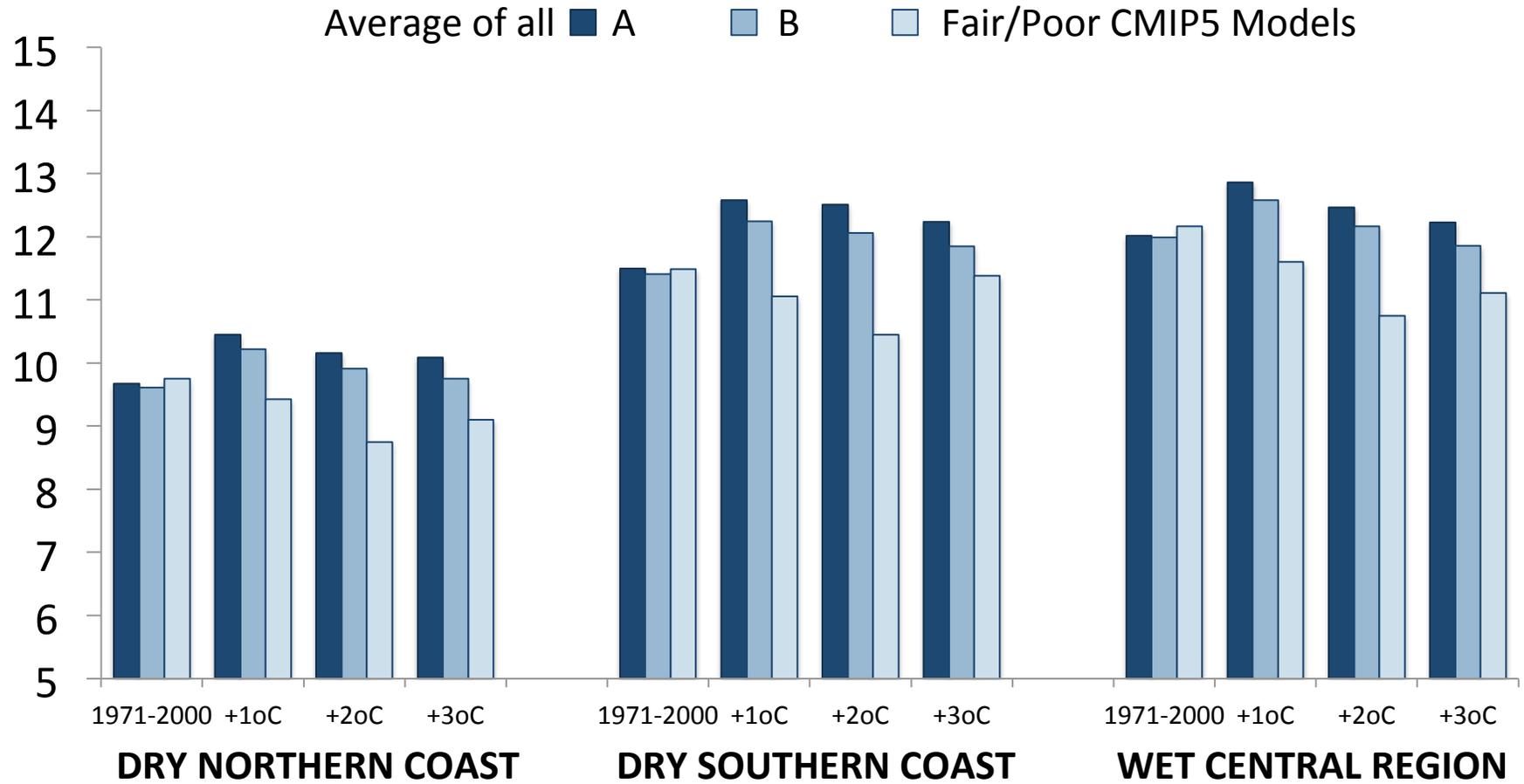
# Precipitation Intensity in Dry Season (mm/day)



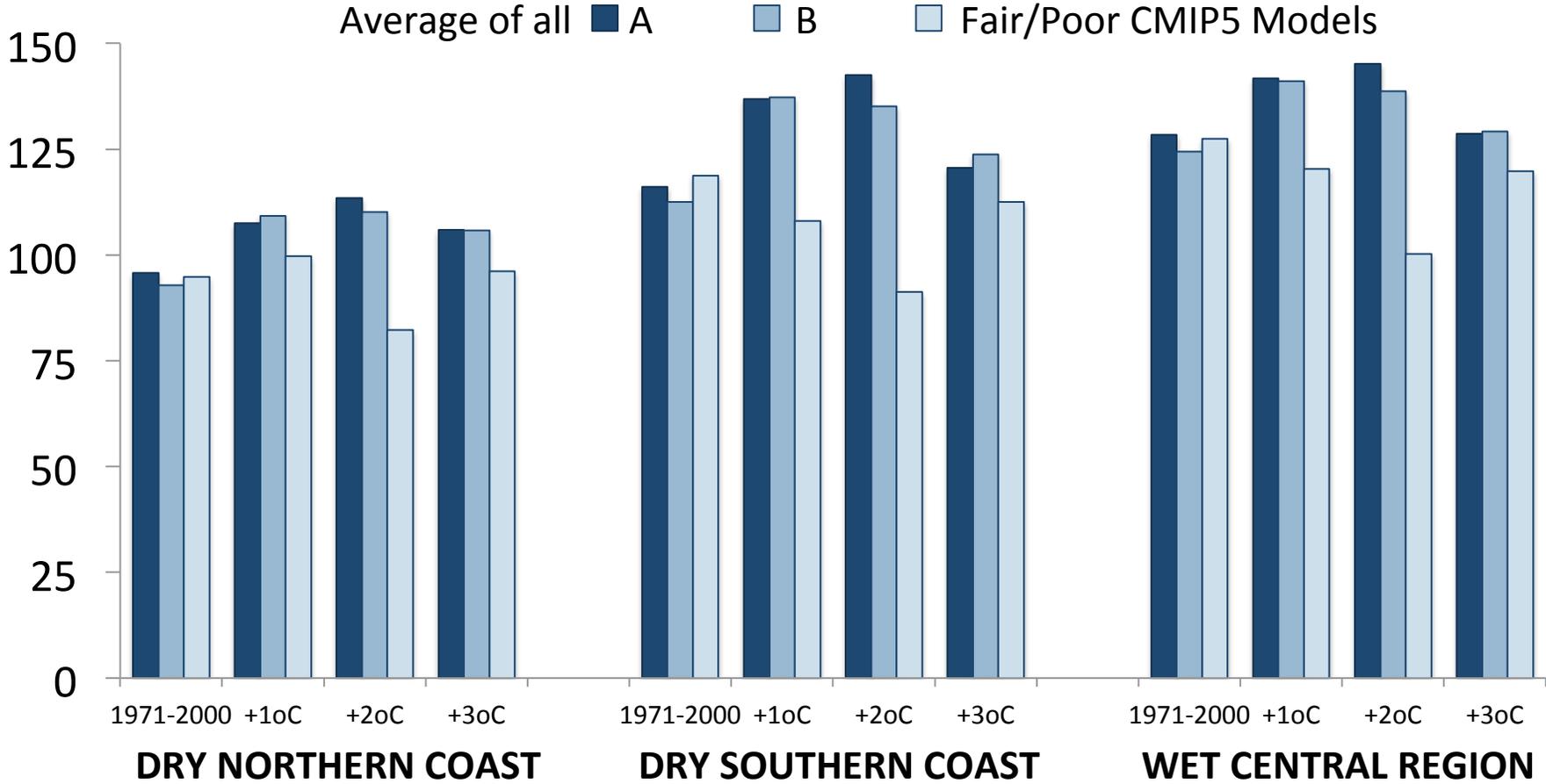
# Precipitation Intensity in Wet Season (mm/day)



# Annual Precipitation Intensity (mm/day)

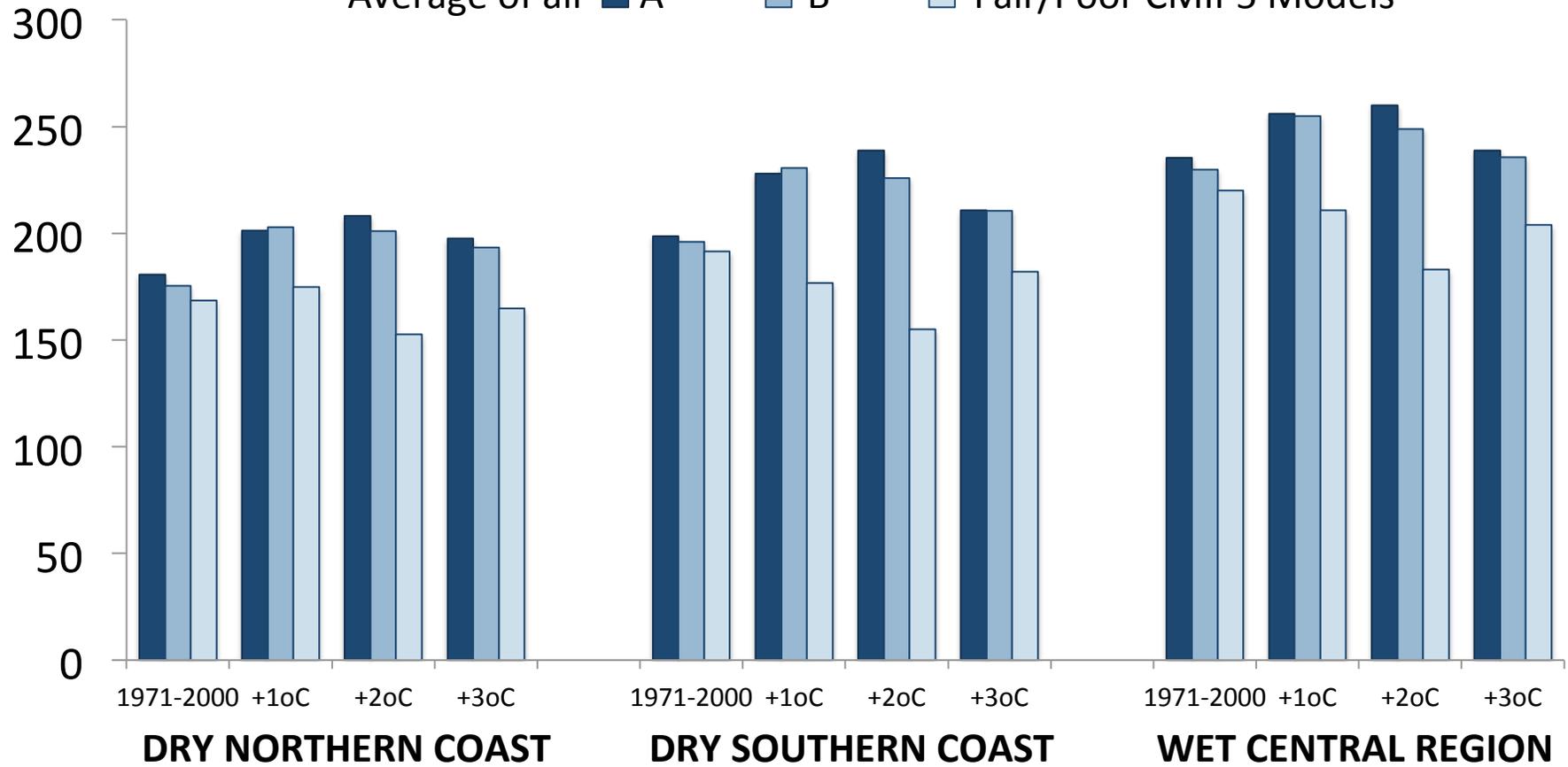


# Precipitation in Wettest 1 Day of the Year (mm)

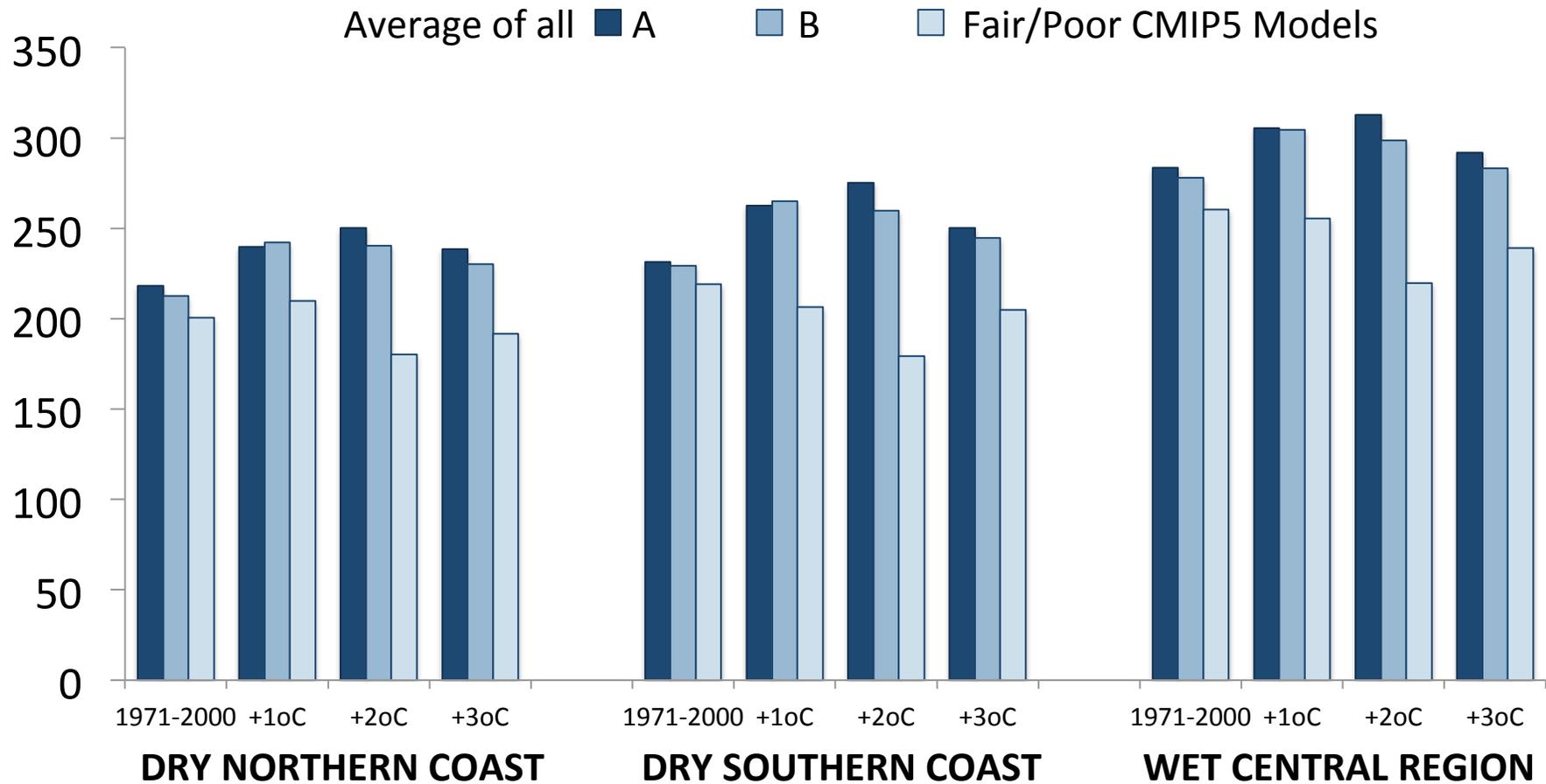


# Precipitation in Wettest 3-Day Period of the Year (mm)

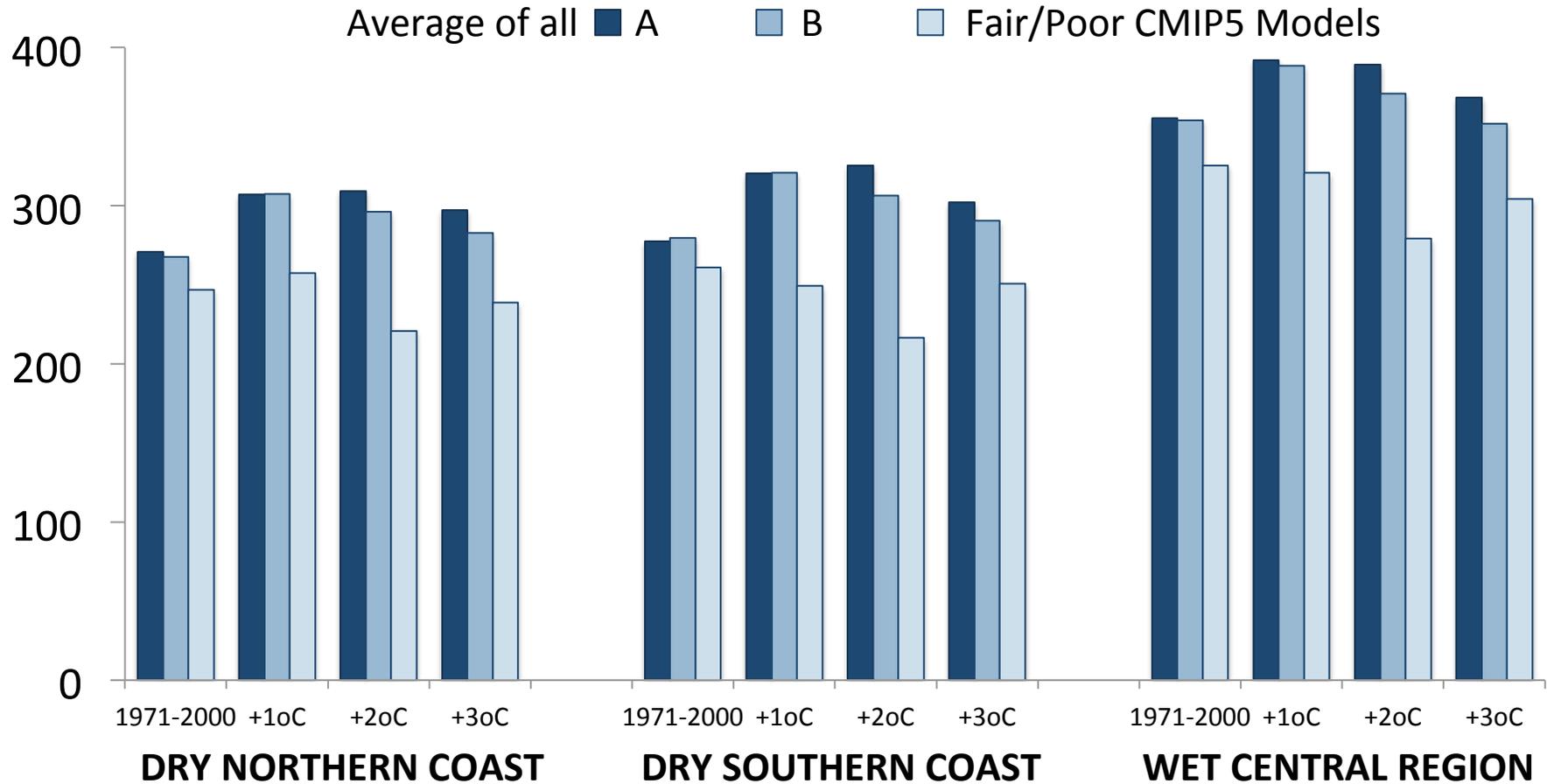
Average of all ■ A ■ B ■ Fair/Poor CMIP5 Models



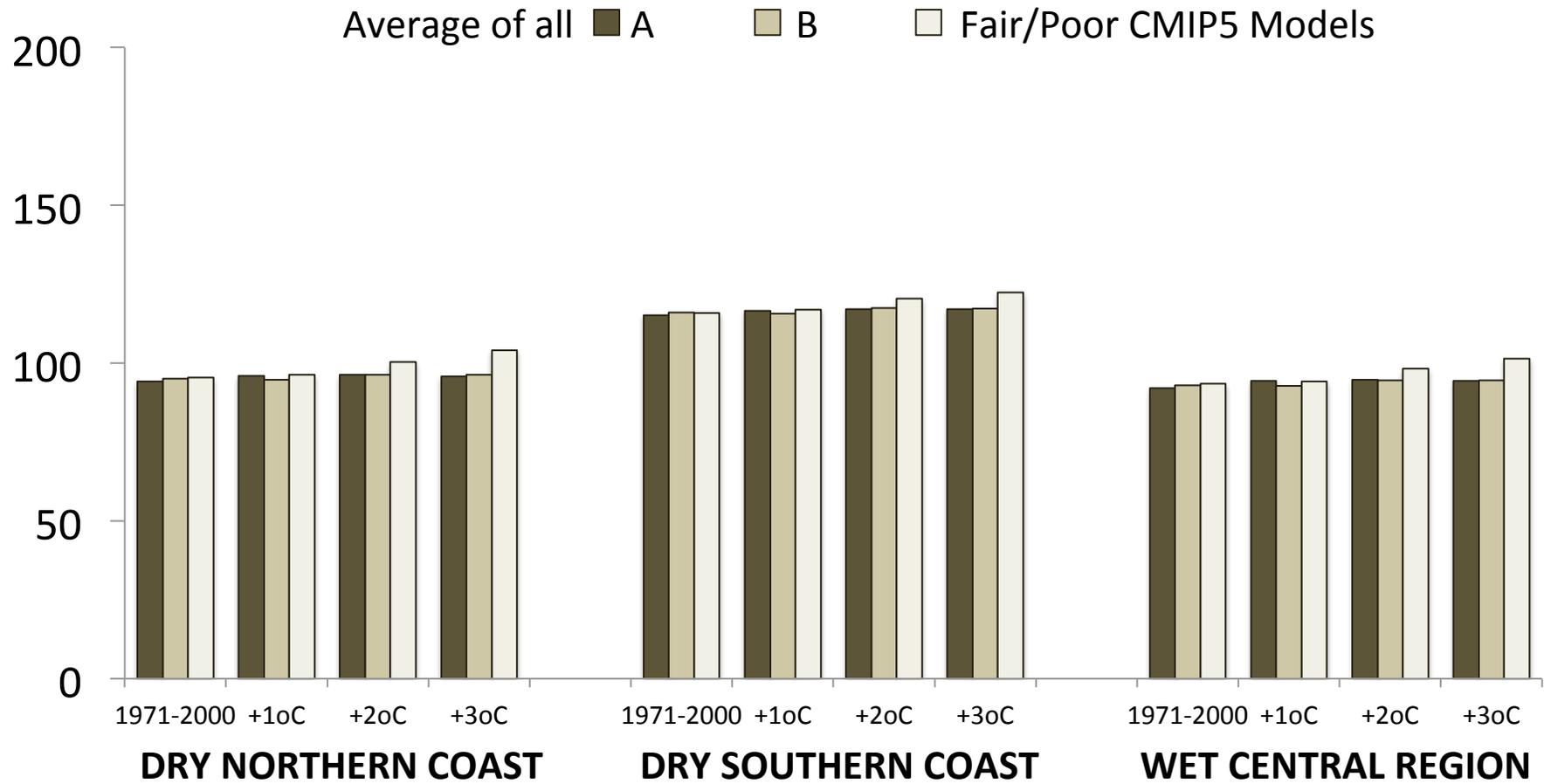
# Precipitation in Wettest 5-Day Period of the Year (mm)



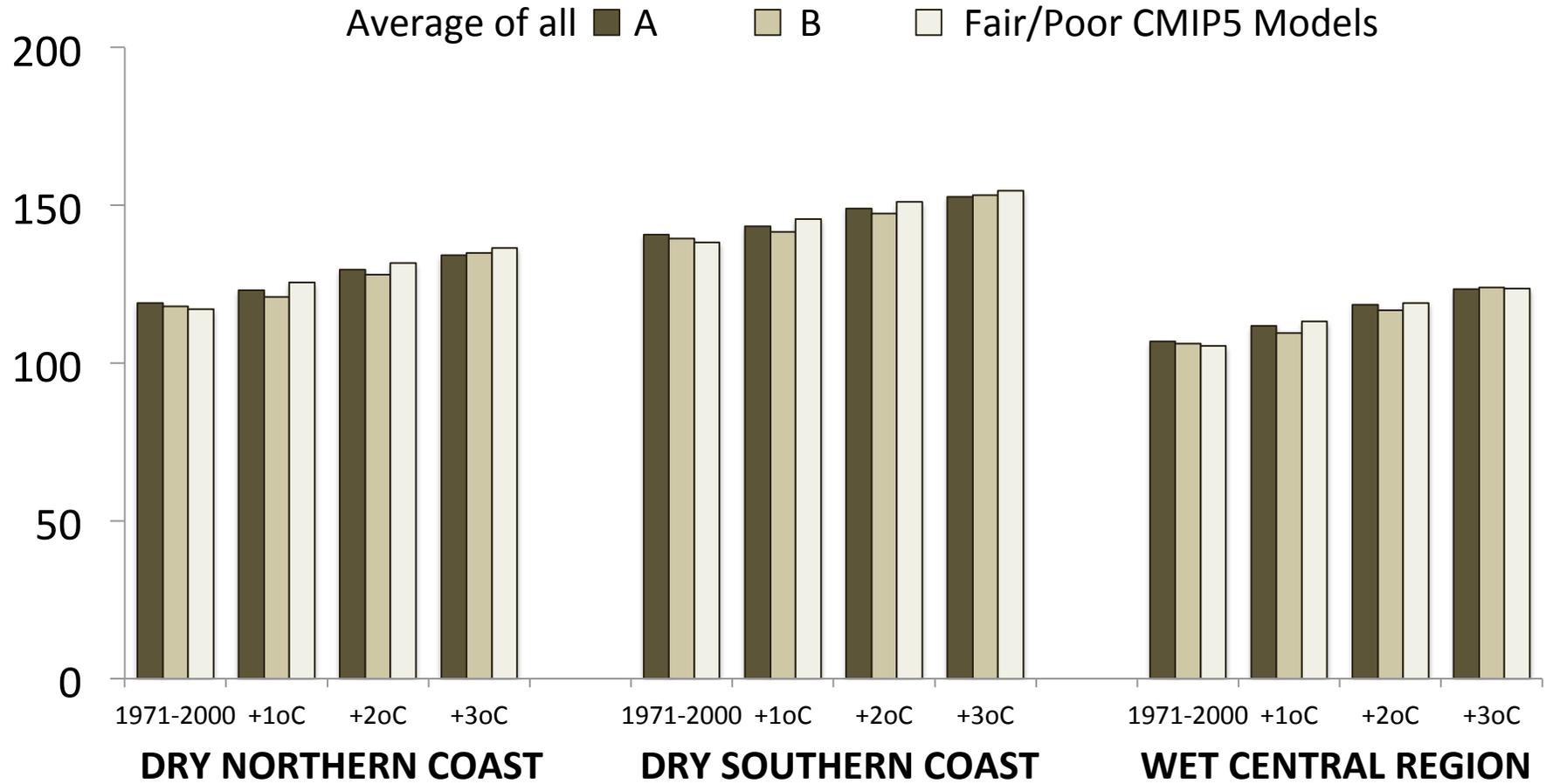
# Precipitation in Wettest 10-Day Period of the Year (mm)



# Number of Dry Days in Dry Season

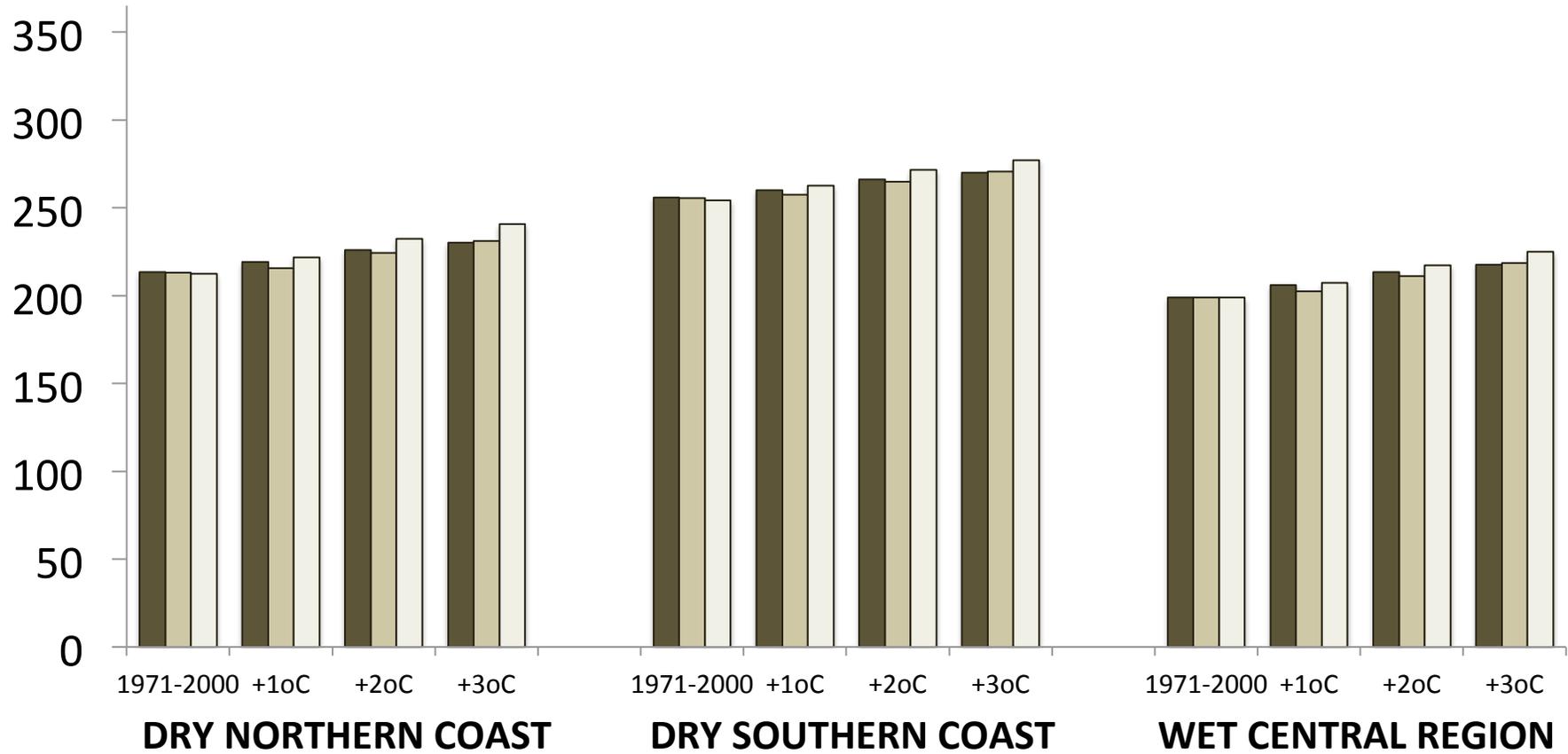


# Number of Dry Days in Wet Season

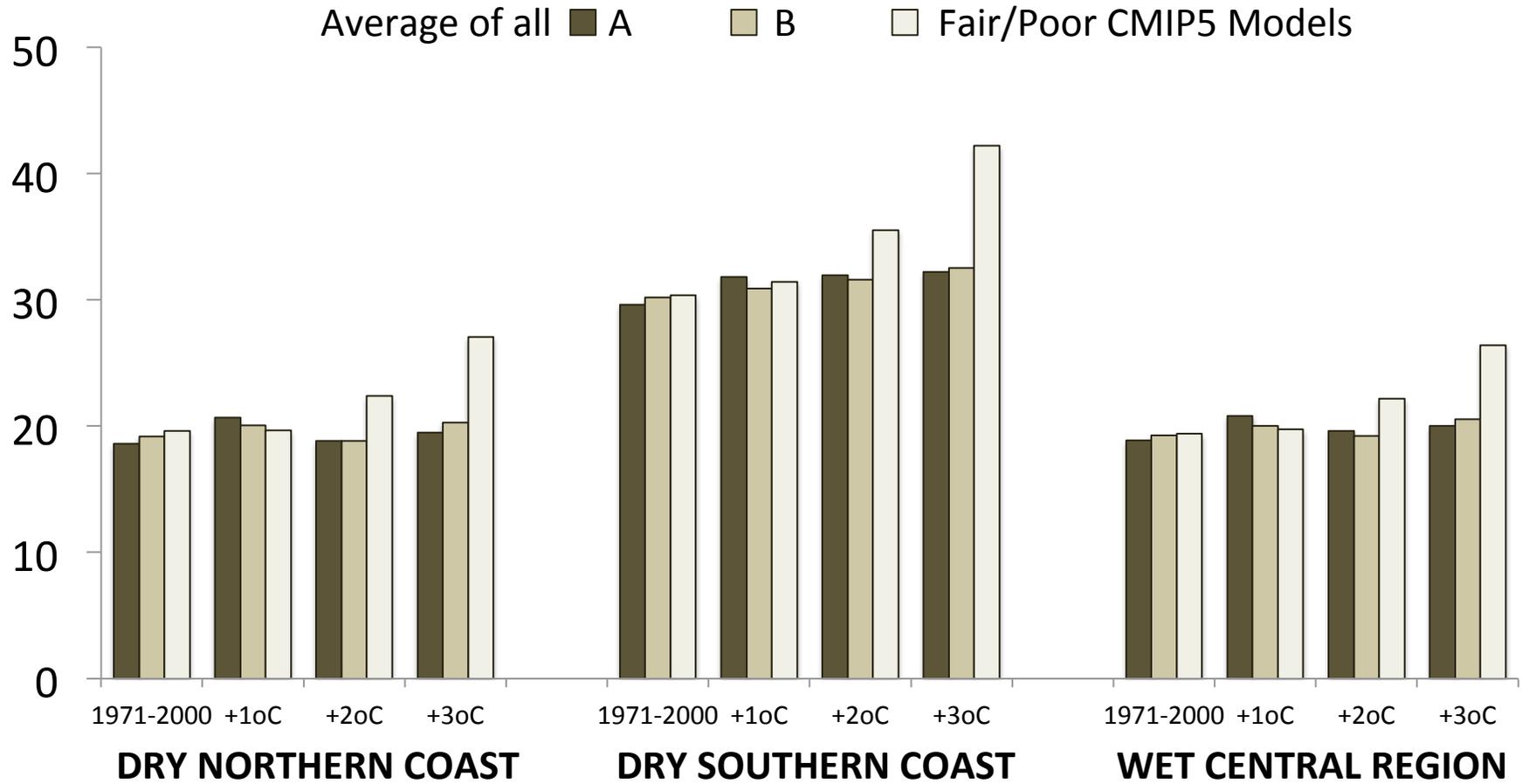


# Number of Dry Days per Year

Average of all **A**    **B**    Fair/Poor CMIP5 Models

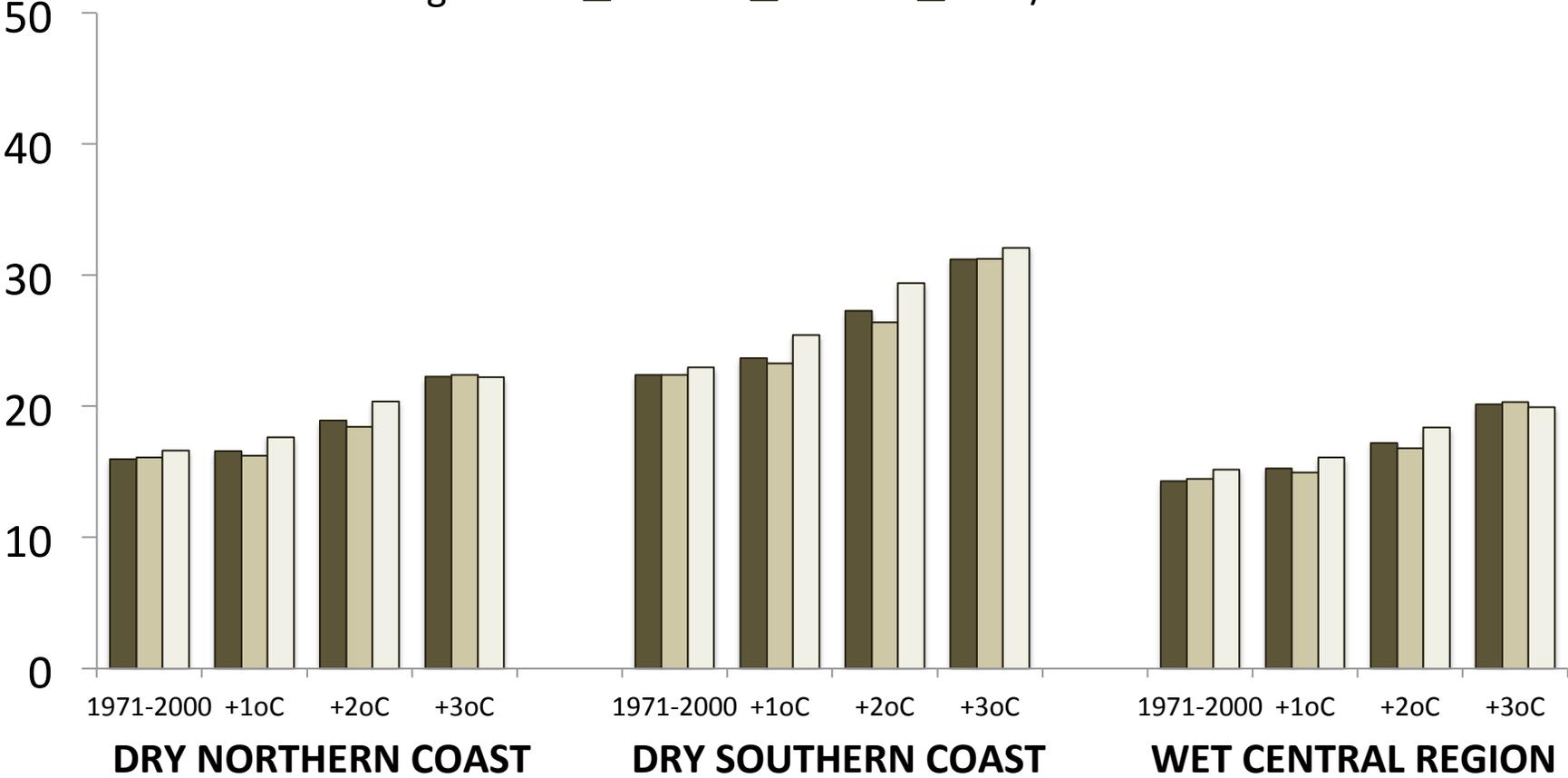


# Longest Dry Period in Dry Season (days)

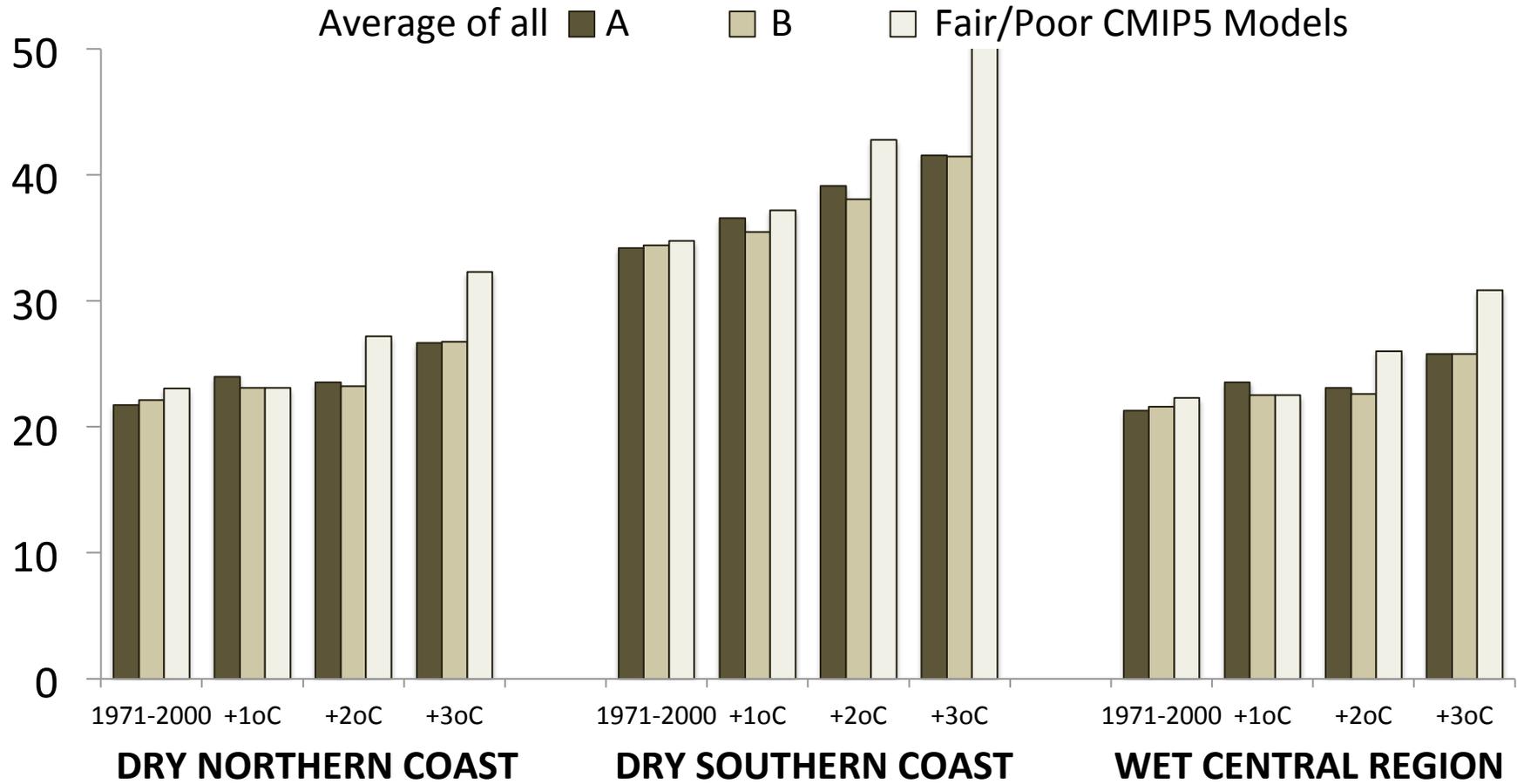


# Longest Dry Period in Wet Season (days)

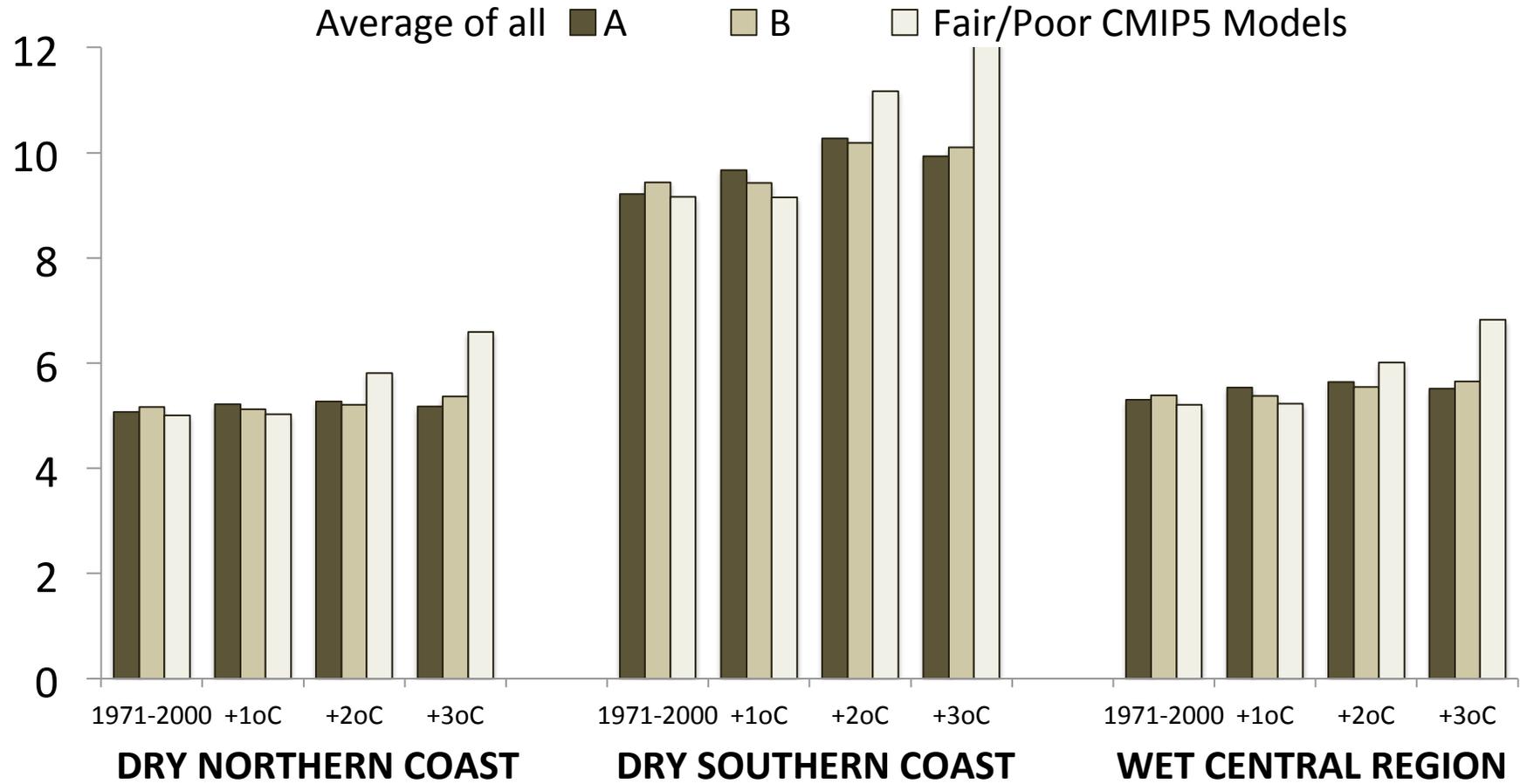
Average of all **A**    **B**    Fair/Poor CMIP5 Models



# Longest Dry Period in Year (days)

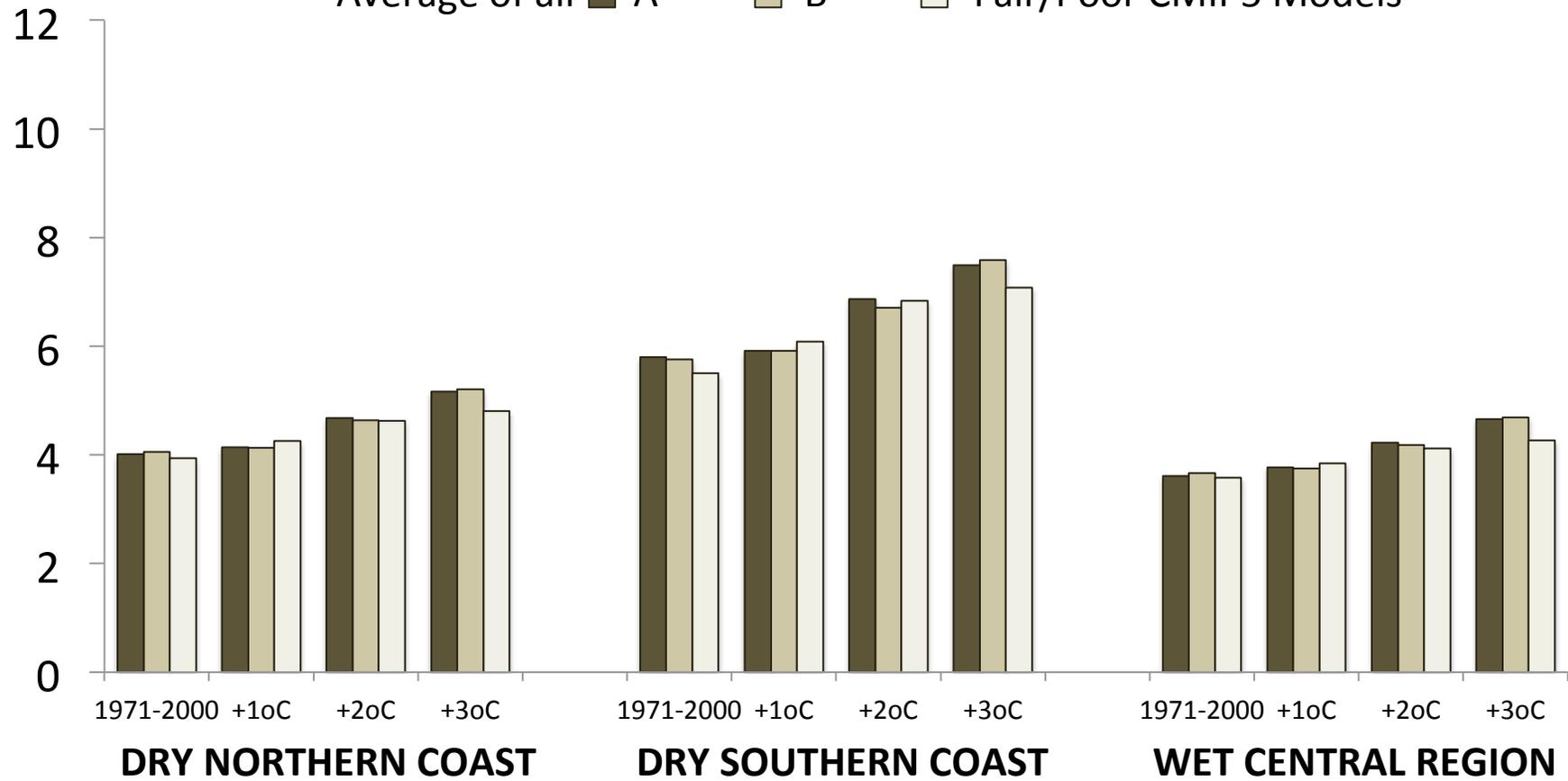


# Average Dry Period in Dry Season (days)



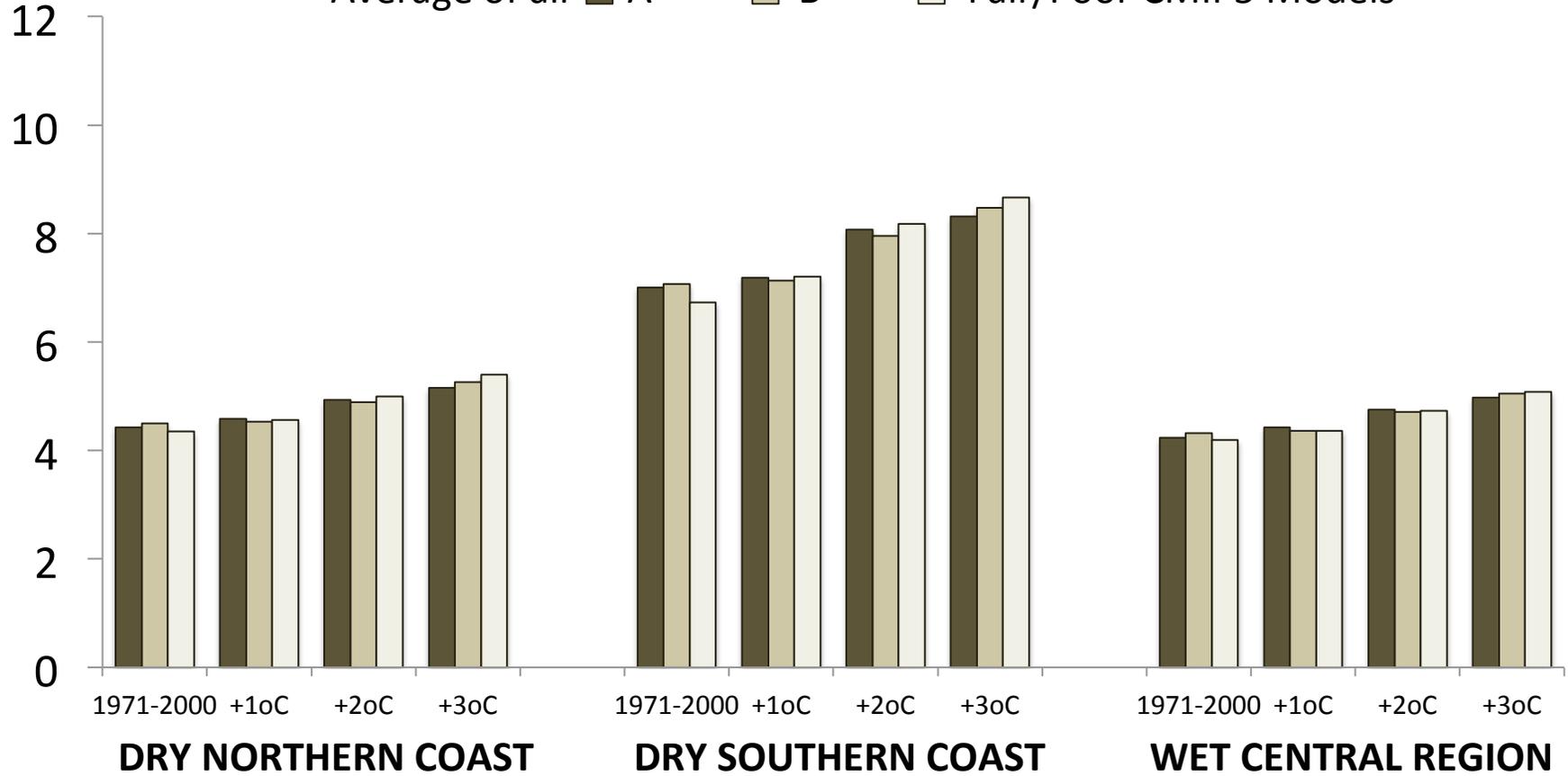
# Average Dry Period in Wet Season (days)

Average of all **A**   **B**   Fair/Poor CMIP5 Models

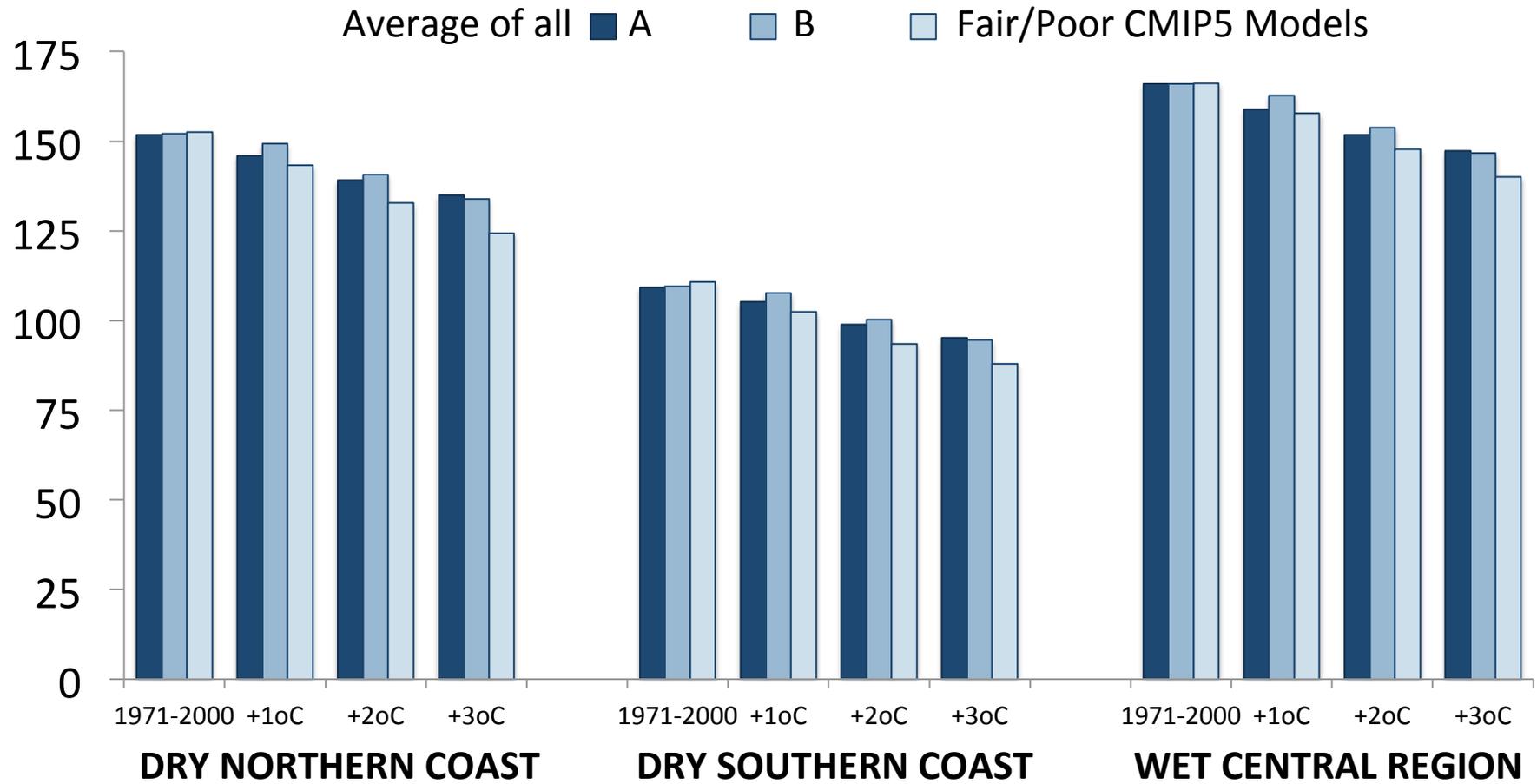


# Average Dry Period in Year (days)

Average of all **A**    **B**    Fair/Poor CMIP5 Models

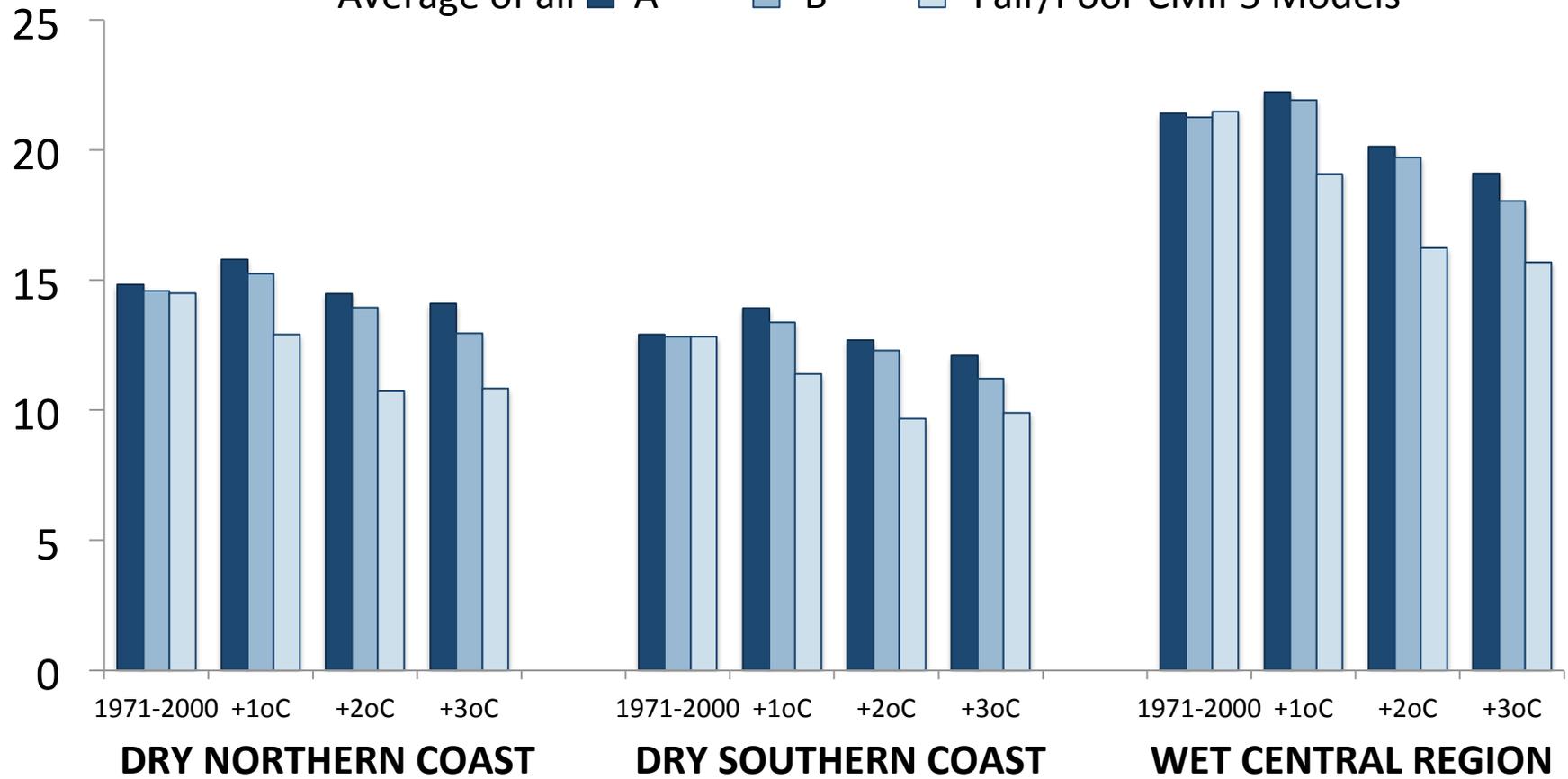


# Number of Wet Days Per Year



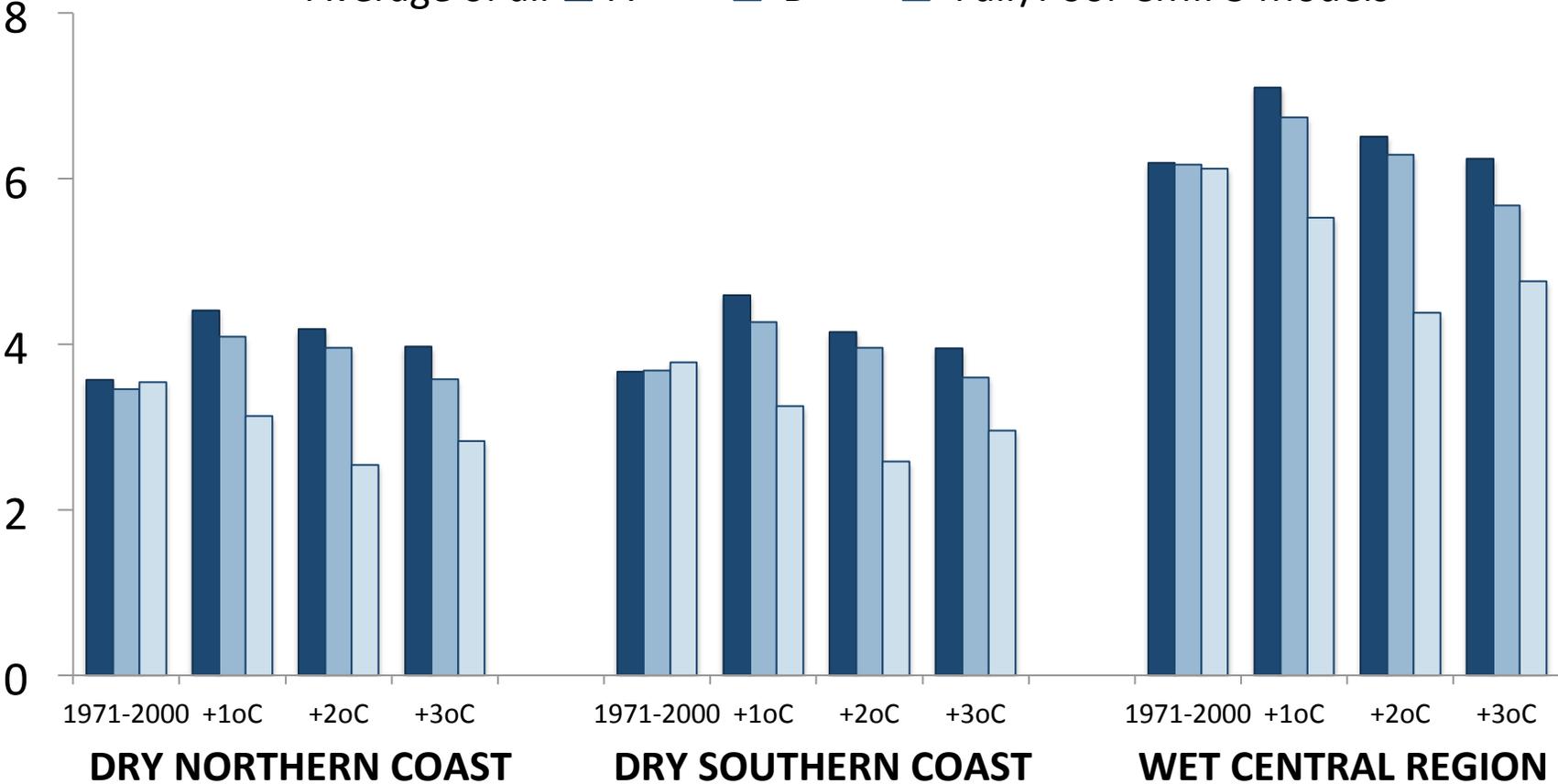
# Precipitation > 1 inch (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



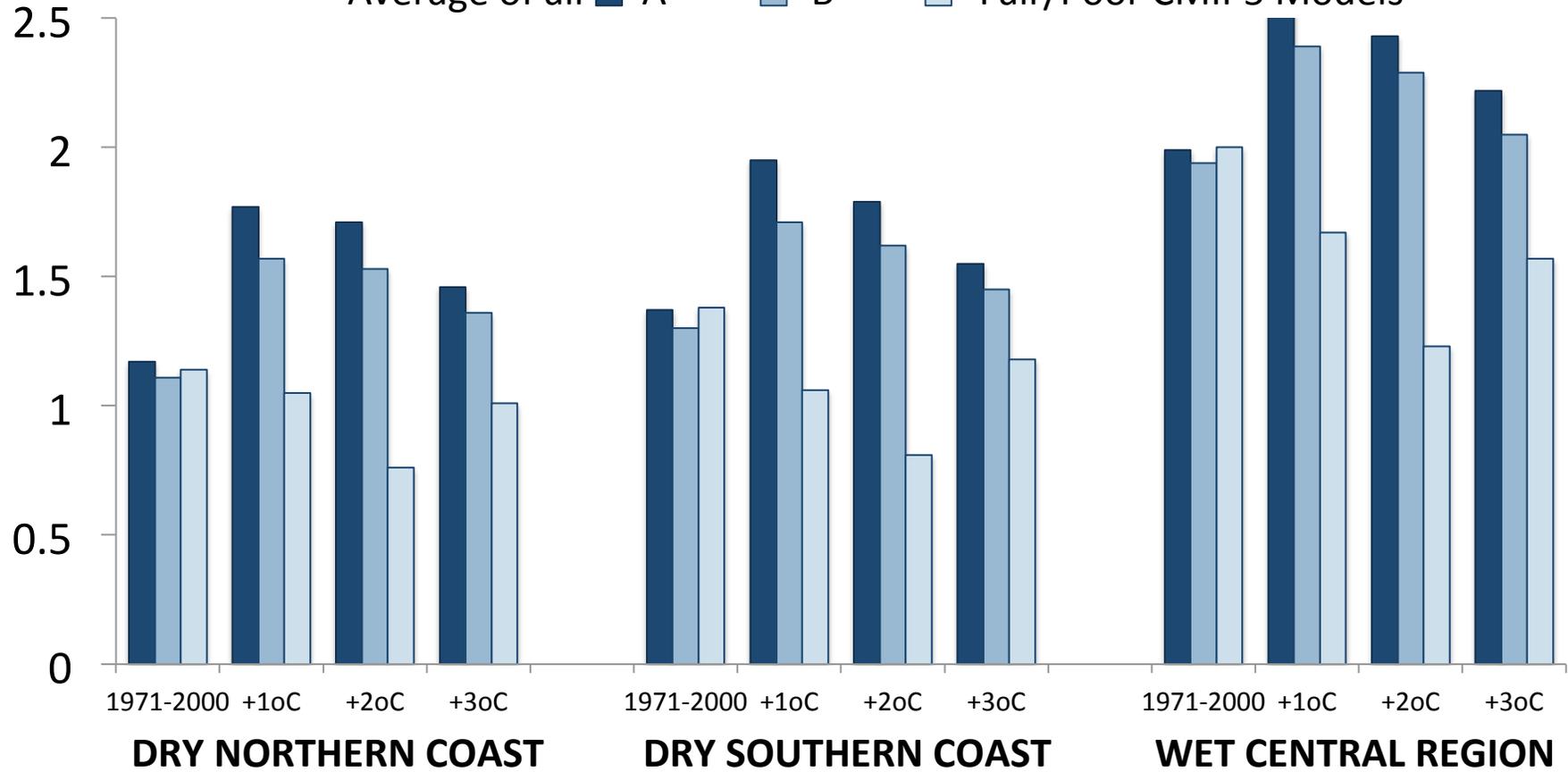
# Precipitation > 2 inches (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



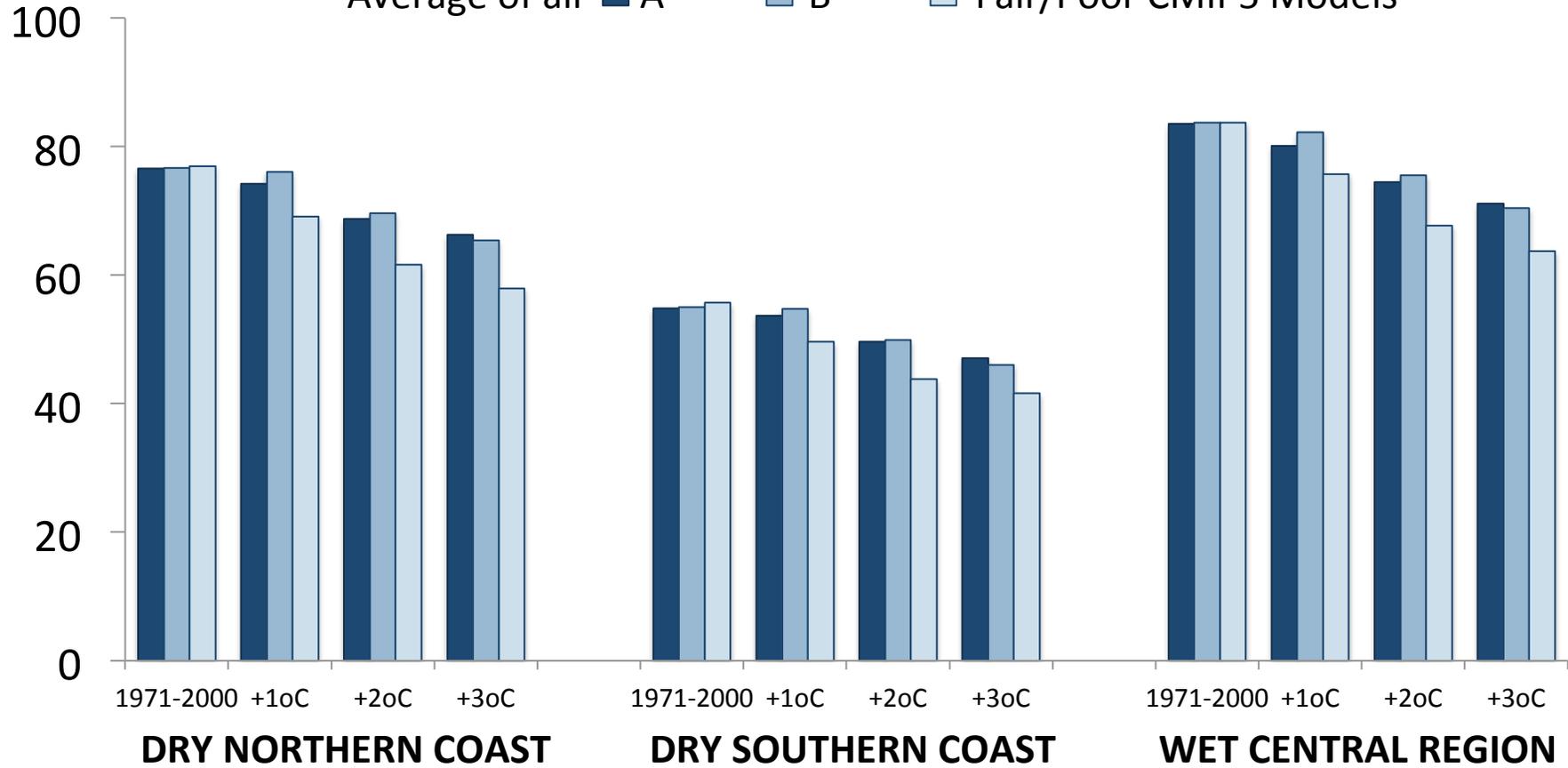
# Precipitation > 3 inches (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



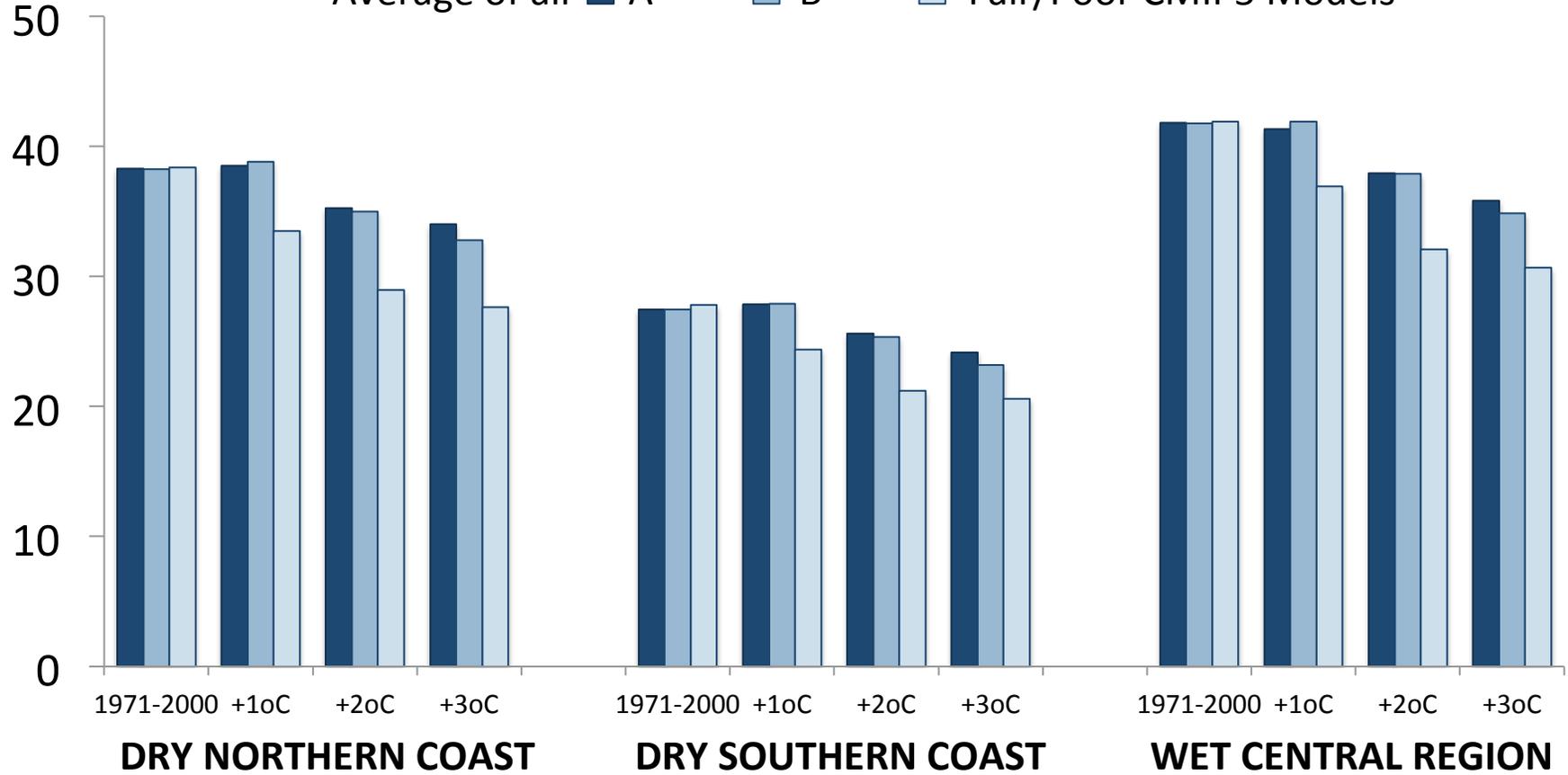
# Precipitation > Historical Median (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



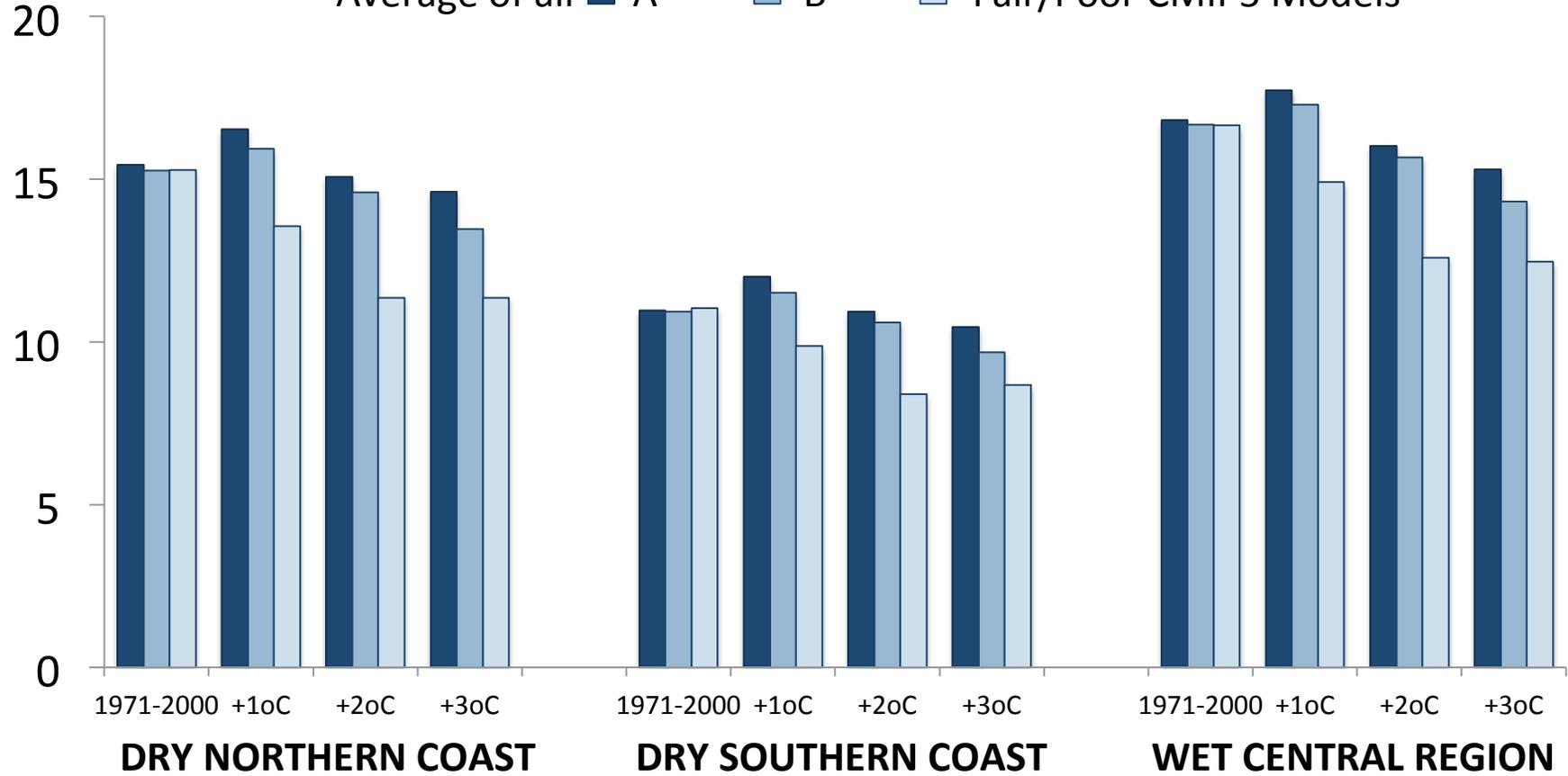
# Precipitation > Historical 75th Percentile (days per year)

Average of all ■ A ■ B ■ Fair/Poor CMIP5 Models



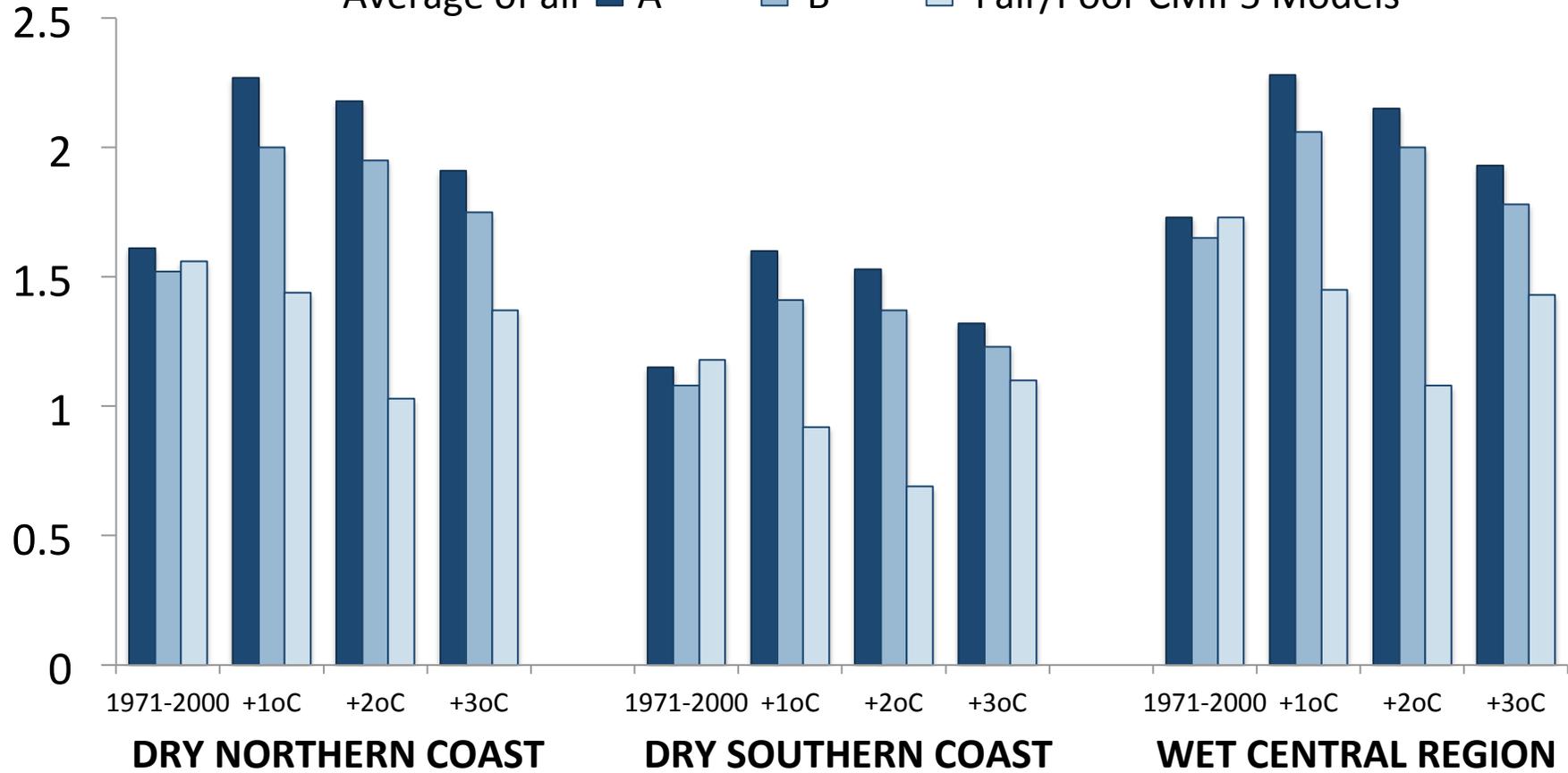
# Precipitation > Historical 90th Percentile (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



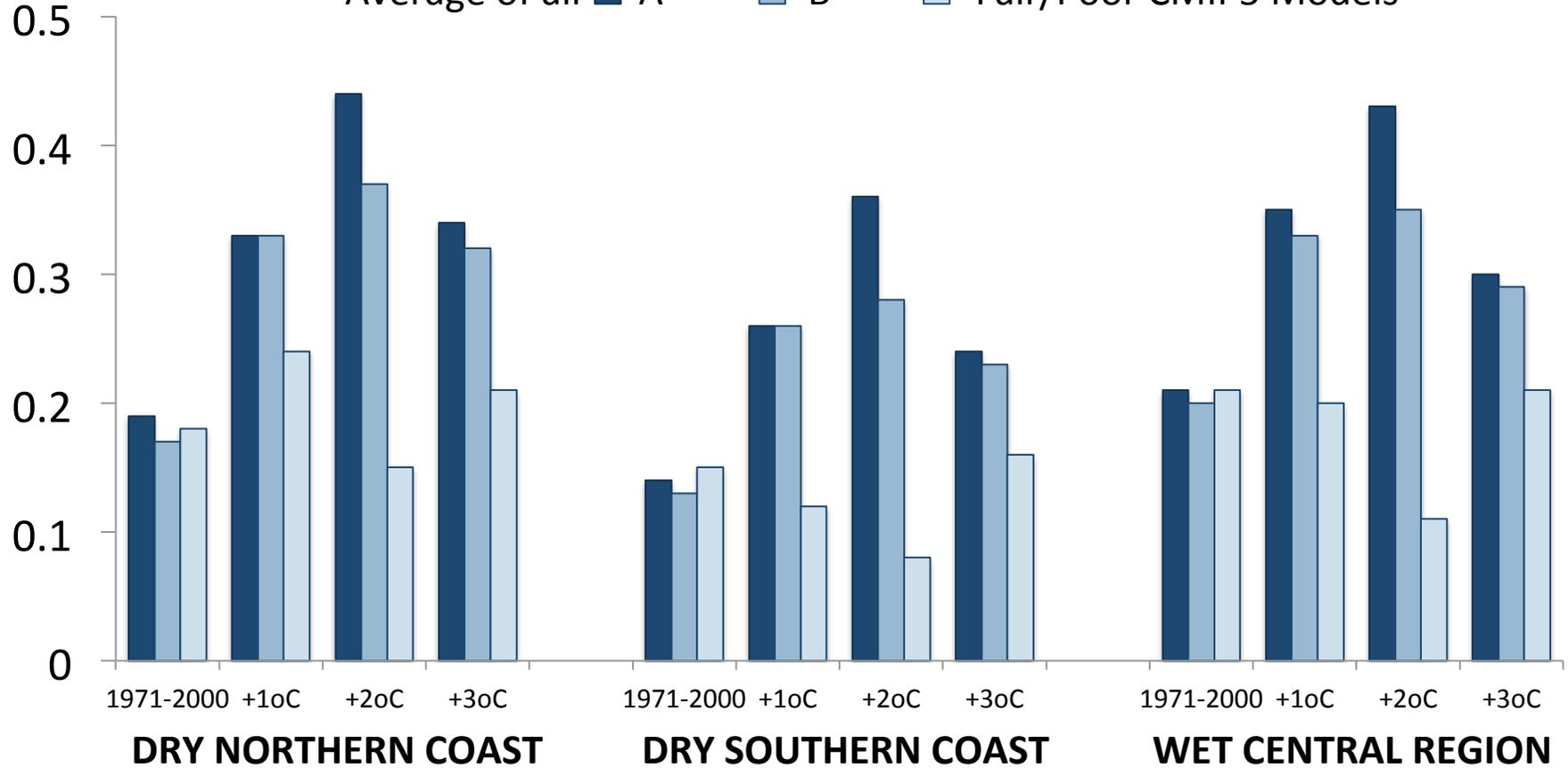
# Precipitation > Historical 99th Percentile (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



# Precipitation > Historical 99.9th Percentile (days per year)

Average of all **A**    **B**    Fair/Poor CMIP5 Models



## **APPENDIX D. A description and evaluation of the statistical downscaling model used to generate high-resolution projections**

Stoner, A., K. Hayhoe, X. Yang and D. Wuebbles (2012) An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*, DOI: 10.1002/joc.3603

# An asynchronous regional regression model for statistical downscaling of daily climate variables

Anne M. K. Stoner,<sup>a\*</sup> Katharine Hayhoe,<sup>a,b,c</sup> Xiaohui Yang<sup>b</sup> and Donald J. Wuebbles<sup>d</sup>

<sup>a</sup> Climate Science Center, Texas Tech University, Lubbock, TX, USA

<sup>b</sup> ATMOS Research & Consulting, Lubbock, TX, USA

<sup>c</sup> Department of Political Science, Texas Tech University, Lubbock, TX, USA

<sup>d</sup> Department of Atmospheric Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

**ABSTRACT:** The asynchronous regional regression model (ARRM) is a flexible and computationally efficient statistical model that can downscale station-based or gridded daily values of any variable that can be transformed into an approximately symmetric distribution and for which a large-scale predictor exists. This technique was developed to bridge the gap between large-scale outputs from atmosphere–ocean general circulation models (AOGCMs) and the fine-scale output required for local and regional climate impact assessments. ARRM uses piecewise regression to quantify the relationship between observed and modelled quantiles and then downscale future projections. Here, we evaluate the performance of three successive versions of the model in downscaling daily minimum and maximum temperature and precipitation for 20 stations in North America from diverse climate zones. Using cross-validation to maximize the independent comparison period, historical downscaled simulations are evaluated relative to observations in terms of three different quantities: the probability distributions, giving a visual image of the skill of each model; root-mean-square errors; and bias in nine quantiles that represent both means and extremes. Successive versions of the model show improved accuracy in simulating extremes, where AOGCMs are often most biased and which are frequently the focus of impact studies. Overall, the quantile regression-based technique is shown to be efficient, robust, and highly generalizable across multiple variables, regions, and climate model inputs. Copyright © 2012 Royal Meteorological Society

**KEY WORDS** statistical downscaling; quantile regression; climate; temperature; precipitation

Received 29 April 2012; Revised 30 August 2012; Accepted 2 September 2012

## 1. Introduction

Atmosphere–ocean general circulation models (AOGCMs) and the new generation of earth system models provide insights into the dynamic nature of possible climate responses to anthropogenic forcing. With spatial scales typically on the order of one half degree or coarser, however, they are unable to simulate climate at the local to regional scale. To compensate for this relatively coarse resolution, a number of dynamical and statistical techniques have been developed to downscale climate model outputs to the impact-relevant spatial and temporal scales at which observations are made.

Despite the plethora of downscaling methods in the literature (Crane and Hewitson, 1998; Wilby *et al.*, 1998; Huth *et al.*, 2001; Stehlik and Bardossy, 2002; Wood *et al.*, 2004; Haylock *et al.*, 2006; Schmidli *et al.*, 2006; Kostopoulou *et al.*, 2007; Hidalgo *et al.*, 2008; to name just a few out of hundreds), relatively few downscaling methods have been applied to quantify potential impacts

of climate change at the local to regional scale for a broad cross-section of regions and sectors across North America. The majority of studies of climate change impacts in the United States, for example, rely on one of five methods: a delta approach whereby a change or ‘delta’ is added to observed mean annual, seasonal, or monthly values in order to get future values (Hay *et al.*, 2000; as used in USGCRP, 2000); simulations from a regional climate model (e.g. Mearns *et al.*, 2009; as used in NARCCAP); the Bias Correction-Statistical Downscaling model originally developed as a front end to the hydrological variable infiltration capacity model, which uses a quantile mapping approach to downscale monthly AOGCM-based temperature and precipitation to a regular grid (Wood *et al.*, 2004; as used in Hayhoe *et al.*, 2004, 2008; Luers *et al.*, 2006; USGCRP, 2009); a constructed analogue approach that matches AOGCM-simulated patterns to historical weather patterns (Hidalgo *et al.*, 2008; as used in Luers *et al.*, 2006); and a linear asynchronous regression approach that downscales daily AOGCM-based temperature and precipitation to individual station locations (Dettinger *et al.*, 2004; as used in Hayhoe *et al.*, 2004, 2008, 2010).

Each of these methods has its own benefits, and each can be sufficient for certain applications. For example,

\* Correspondence to: A. M. K. Stoner, Climate Science Center, Texas Tech University, 113 Holden Hall, Boston & Akron Streets, Lubbock, TX 79409-1015, USA. E-mail: anne.stoner@ttu.edu

the simple and transparent delta approach can yield a nearly identical downscaled annual or seasonal mean temperature value as a more complex statistical model. At the other end of the spectrum, complex regional climate models are computationally demanding, but provide consistent high-resolution projections for a plethora of surface and upper-air variables. None of these five methods, however, allows for using multiple climate models and scenarios as input while downscaling to any spatial scale (including both station-based and gridded), simulating additional impact-relevant variables (such as solar radiation and humidity), and adequately resolving projected changes in daily climate extremes, at the same time.

For that reason, we have developed a new statistical downscaling model, the asynchronous regional regression model (ARRM). ARRM builds on the same statistical technique used by the last downscaling approach listed above (Dettinger *et al.*, 2004), asynchronous quantile regression, to define a quantitative relationship between any daily observed and simulated surface variable that has a symmetric distribution, with particular emphasis on accurately resolving the relationship at the tails of the distribution in order to capture simulated changes in extremes. Asynchronous quantile regression removes the time stamp from historical observations and simulations, reordering each time series by value before matching quantiles of observed data with those from AOGCM output. This is important because coupled AOGCM simulations generate their own patterns of natural variability, meaning that no day-to-day or even year-to-year correspondence with observations should be expected.

The general concept of quantile regression was originally developed in the field of econometrics by Koenker and Bassett (1978) to estimate conditional quantiles of the response variable as opposed to the conditional mean estimated by the orthodox least-squares regression method. The quantile regression approach is of particular utility to geospatial data, in that it can be used to determine relationships between two quantities that are not measured simultaneously, such as an observed and a model-simulated time series. It takes advantage of the hypothesis that although the two time series may be independent, their distributions may be similar.

The general technique of quantile regression has been used in a variety of applications, including by O'Brien *et al.* (2001) to determine relationships between measurements of relativistic electron conditions measured from two different satellites passing over the same area at different times. Dettinger *et al.* (2004) were the first to apply this method to downscaling AOGCM output, to examine simulated hydrologic responses to climate change. In this application, the first time series was observations and the second, historical model simulations. The regression model derived from these two distributions was then applied to transform the distribution of, or downscale, future model simulations.

The objective of this study is to build on the foundation of quantile regression to develop a relatively straightforward, flexible, efficient, and robust statistical model

that is capable of downscaling any atmospheric variable, measured on a daily or monthly basis, which has, or can be transformed into, an approximately symmetric distribution. Section 2 describes the statistical basis of the model and refinements that improve its ability to downscale global model outputs. Section 3 describes the long-term weather station observations and the AOGCM outputs used to evaluate the downscaling model in terms of its ability to simulate observed temperature and precipitation, using the same variables from the AOGCMs as predictors. Section 4 describes how the model was developed in multiple steps, each of which is successively tested to ensure that the additions improve the model's ability to reproduce historical climate. Section 5 discusses the results of applying the downscaling model to end-of-the-century temperatures and precipitation and the changes between downscaled and raw AOGCM output compared with present conditions. Finally, Section 6 summarizes the findings of this study.

## 2. Model development

### 2.1. Statistical basis

The concept of *quantile regression* was first introduced by Koenker and Bassett (1978), where quantiles refer to values of a cumulative population (i.e. when the data are sorted by increasing value) that divide the population into equal-sized segments. Quantiles are the data values marking the boundaries between consecutive subsets. If the data are divided into  $q$  equal-sized subsets, the  $k$ th quantile for a variable is the value  $x$  such that the probability that the variable will be less than  $x$  is no greater than  $k/q$  and the probability that the variable will be more than  $x$  is no greater than  $(q - k)/q$ . A distribution has  $q - 1$  quantiles, one for each integer  $k$  satisfying  $0 < k < q$ .

In general, regression analysis quantifies covariance between variables, and, if it exists, provides a model to predict one variable on the basis of the other variables used as input to the regression. Quantile regression specifically estimates *conditional quantile functions* – models in which quantiles of the distribution of the predictor variable are expressed as functions of observed covariates (Koenker and Hallock, 2001). In other words, quantile regression results in estimates approximating the quantiles of the predictor variable. For a time series containing  $N$  values there are  $N$  ranks in each vector. A model can be constructed by regressing the value at rank  $n_i$  of the simulated vector onto the value of the same rank of the vector containing observed values, for  $i = 1 \dots N$  (as done for example in Dettinger *et al.*, 2004). This regression is asynchronous, i.e. data values that are regressed against each other did not necessarily occur the same calendar day, but rather correspond by quantile or rank. The regression model derived from historical AOGCM simulations and historical observations can then be applied to future AOGCM simulations, to project downscaled future conditions.

Asynchronous regression is an important component of this model, because a coupled AOGCM simulation is free to evolve chaotically, with only the external forcings being prescribed; hence, each simulation represents one out of many possible outcomes and no daily correspondence between the model and observations should be expected.

## 2.2. Model input

Both theoretical and practical considerations affect the selection of inputs to quantile regression. First, it is important to verify that the two time series (simulated and observed, or predictor and predictand) have somewhat similar distributions; the closer both distributions are to Gaussian, the simpler the function relating the two distributions. Even non-Gaussian distributions can sometimes be manipulated to mimic a Gaussian distribution; here, in the case of precipitation, by taking the natural logarithm of the daily wet day precipitation values.

To train the downscaling model, the observed record must have an adequate length and quality of data. A minimum of 20 consecutive years of daily observations with less than 5% missing data is usually needed in order to appropriately sample from the range of natural climate variability at most of the station locations examined here and to produce robust results without overfitting. To challenge the downscaling model, two stations were selected for this evaluation that had substantially less data available (Bridgeport, WV with 78% and Moosehead Lake, ME with 88% of daily data missing over 50 years).

## 2.3. Model structure

The structure of the ARRM model is summarized in Figure 1. The first step is to prepare the data by separating it into 12 vectors by month such that a separate statistical model can be built for each month. This accounts for different weather patterns dominating any given region at different times of the year that could alter AOGCM biases relative to observations. Two weeks of overlapping data on either side of each month are included to account for future conditions that may lie outside the range of a typical historical month. This extension also doubles the use of each data point during the training process. Each month's time series is then reordered by rank to create an asynchronous vector. Figure 2 shows AOGCM-simulated (grid cell containing the weather station) versus observed temperature for chronological and for sorted data, illustrating how ranking of the inputs provides a correlation between observations and model simulations whereas matching by calendar date does not.

The second step in the ARRM model is to fit a regression function to the ranked values shown in Figure 2(b). For most station locations and global models, a linear fit (as used in Dettinger *et al.*, 2004) is adequate within at least the 20th–80th percentiles of the distribution (dark-coloured line in Figure 3) with a high coefficient of determination ( $R^2$ ). However, residuals are often large

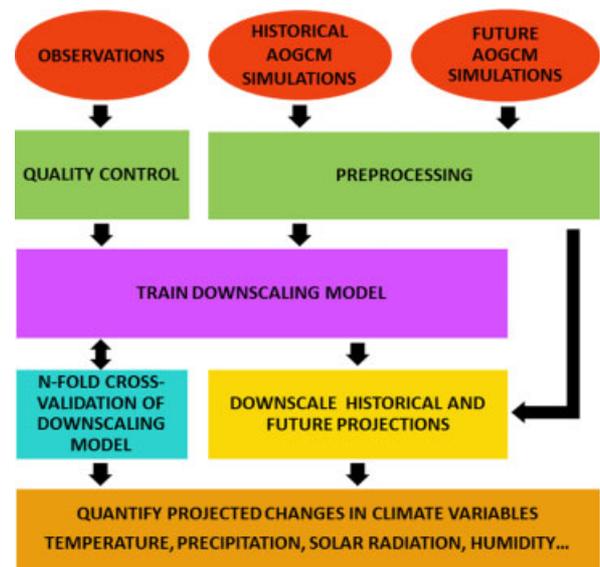


Figure 1. Structure of the ARRM downscaling model. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

near the tails of the distribution that, depending on the application, can be of greater interest to climate impact studies than values at the centre of the distribution. Polynomials of increasing order result in increasingly better fits to the historical observations (not shown), but run two serious risks: first, of overfitting, and second, of exhibiting unnatural behaviour at the tails of the distribution that could unrealistically predict lower observed temperatures for higher modelled values than for lower modelled values, and vice versa.

Instead, we found that a piecewise linear regression (light-coloured, segmented line, Figure 3) provided the most consistent fit while accounting for biases in model values near the tails of the distribution; biases that can be markedly different than those simulated for values near the centre of the distribution. Adding breakpoints, or knots, allows for different slopes at different parts of the distribution, in particular minimizing the residuals at the tails of the distribution when compared with either a linear or a polynomial fit.

*R* (R Development Core Team, 2012), the statistical programming language used to build ARRM, has spline-based functions such as *bs* and *ns* that can add breakpoints to a regression. However, these functions require the user to set the number of breakpoints manually and then place the points at predetermined, evenly distributed, quantiles. As illustrated in Figure 3, the ideal number of breakpoints can vary broadly, depending on the characteristics of model bias for a given month and/or location. A new function was therefore required that would optimize the regression model for each month by automatically identifying the number and location of up to six independent breakpoints. This piecewise linear regression function is described next.

The third step in the ARRM model is to use the statistical regression models, constructed from observed and historical simulated time series, to downscale future

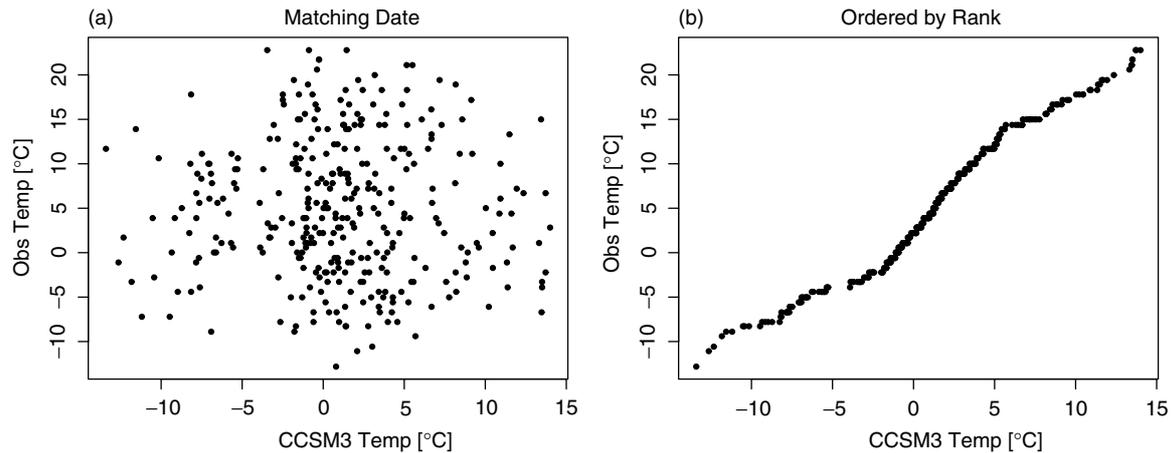


Figure 2. Monthly scatter plot of observed and AOGCM (CCSM3) simulated (nearest grid cell) daily maximum 1960–2009 January temperature for Bridgeport, West Virginia, matched (a) by calendar date and (b) by rank.

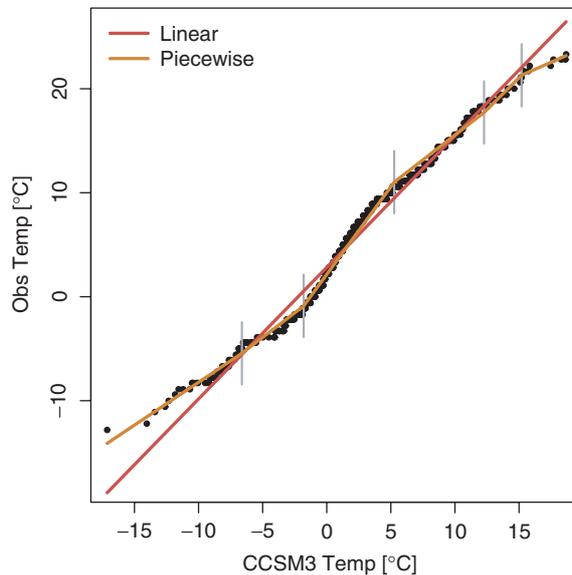


Figure 3. Scatter plot of observed *versus* AOGCM (CCSM3) simulated daily maximum 1960–2009 January temperature for Bridgeport, West Virginia (including half of previous and following months), ranked by quantile. Dark-coloured line shows the results of a linear regression on the data; lighter, segmented line shows the results of a piecewise linear regression; and vertical grey lines identify knots in the piecewise linear regression. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

projections. The resulting downscaled values must subsequently be rearranged back into the original order to retrieve the final product, a continuous chronological time series of the downscaled values.

#### 2.4. Piecewise linear regression function

The piecewise linear regression function developed for ARRM is based on linear regression that iterates over a moving window. For the majority of the distribution, the window width remains fixed at a given percentage of the total number of data points for that particular month. As the concentration of data points near the tails of the distribution is much sparser than at the centre, window

width at the tails of the distribution decreases linearly to a minimum width by the ends of the distribution.

This function requires four fixed settings: the percentage of data points in the fixed window width, the minimum and maximum probabilities over which a fixed window width is used, the minimum permissible width of the window at the tails of the distribution, and the maximum number of breakpoints allowed. Optimal values for these settings are a function of AOGCM bias, characteristics of which differ from one variable to another. In general, a fixed window width of 5% of the distribution between probabilities of  $0.1 < P < 0.9$ , linearly decreasing to a minimum width of either  $2^{\circ}\text{C}$  or ten data points (whichever is greater) for  $P < 0.1$  and  $P > 0.9$ , is adequate for temperature as the relationship between observed and modelled values tends to be relatively linear over much of the distribution. For precipitation, greater variability in AOGCM bias over the distribution requires a wider fixed window width, on the order of 10%, between probabilities of  $0.15 < P < 0.85$ , linearly decreasing to a minimum of 5% of the mean value or ten data points (again, whichever is greater) for  $P < 0.15$  and  $P > 0.85$ .

Up to six breakpoints are allowed in each regression model. This number was determined on the basis of two factors: first, visual testing by plotting downscaled projections for the historical period for individual months showed that more breakpoints tended to increase the risk of overfitting, such as introducing shorter segments with negative slopes, particularly for months with sparse data or poor model performance, and second, months with dense data rarely required more than six breakpoints and often far less. The function begins the piecewise regression at the 40th percentile, where the data point at the 40th percentile is the largest value in the window and moves up (to the highest quantile of the distribution) from there. In other words, the moving window starts with the  $X\%$  data points below the 40th percentile, where  $X$  equals 5 for temperature and 10 for precipitation. The selection of the 40th percentile is to ensure that

the middle part of the distribution is well covered by a moving window. QR matrix decomposition is used to fit a linear regression to the data in the window. The  $R^2$  value for each regression is recorded and saved in a vector, and the moving window is shifted up one data point towards the end of the distribution until it reaches the 100th percentile. The first breakpoint is defined as the central point of the window with the lowest  $R^2$  value of the vector, if the value of  $R^2$  for that window is less than the value for the entire time series. The  $R^2$  values on either side of that breakpoint are then blocked for the width of the moving window and a new minimum identified, for a total of up to three breakpoints in the upper half of the distribution.

This process is repeated beginning at the lowest found breakpoint, or if no breakpoints are found, at the 40th percentile moving down to the 0th percentile. This time, the moving window trails above the percentile. This allows an  $R^2$  value to be assigned to each data point in the monthly vector, from the first to the last. Setting a minimum window width of ten points means that breakpoints are not allowed to fall within the first and last five points of the dataset.

Before the statistical model is finalized, slopes between breakpoints are automatically reviewed. Breakpoints that create a negative slope can cause lower AOGCM values to produce higher downscaled values than higher AOGCM values. Breakpoints that create a slope close to zero ( $-0.1 < \text{slope} < 0.1$ ) can create an unrealistic peak of nearly identical values in the downscaled distribution. Removal of a breakpoint causing a negative or 'flat' slope will always have a detrimental effect on the  $R^2$  value of the regression fit, because the segment having the negative or 'flat' slope yielded the best fit, but improve the realism and generalizability of the fit. Sometimes, when AOGCM biases are particularly nonlinear, the removal of negative slopes can have a greater impact on the quality of the fit than the impact of having a few data points with downscaled values that decrease rather than increase for a small interval within the distribution. Hence, the function allows for negative or 'flat' slopes under two conditions: if they are not the first or last segment in the regression, and if they span less than five points. If these conditions are not met, the breakpoint below the negative slope is removed unless it is the first segment of the regression, in which case the breakpoint to the right is removed. One breakpoint is removed at a time and the process repeated once the regression and new slopes have been recalculated to determine whether a new segment with a negative slope has been introduced. This process is repeated until all negative or flat slopes have been eliminated.

Once the breakpoints have been finalized, the regressions are used to build a statistical model that performs piecewise linear regressions, with the use of spline interpolations, between the monthly simulated and observed data ordered by rank. This regression model can then be used to downscale future values, similarly ordered by rank, assuming stationarity in climate system feedback mechanisms.

## 2.5. Bias correction

As ARRMs are statistical models, there is a risk of introducing unrealistic values especially at the tails of the distribution, where data points are sparse and the slope of the initial and/or final regression can be very sensitive to a single extreme point. In some cases, an observational data point may even be in error. An example is the Global Historical Climatology Network (GHCN) dataset for Hialeah, FL, which had a recorded maximum temperature for 8 November 2003 with a value of  $-17.8^\circ\text{C}$ ,  $25^\circ\text{C}$  lower than the second lowest maximum temperature recorded for this station, and with temperatures for the previous and following days of  $29.4$  and  $30.0^\circ\text{C}$ , respectively. This erroneous point noticeably affected the magnitude of predicted cold temperature extremes for this location. Unrealistic values in the original observations are therefore removed by the quality control procedure described in Section 3, prior to their use as input to the downscaling model.

Because of this sensitivity, downscaled extremes (defined as lying below the 5th percentile and/or above the 95th percentile of the distribution) that fall outside a realistic range for each station are corrected separately, by calculating the bias in percent difference between the downscaled value and the minimum or maximum observed value for that location. To avoid large biases that can be caused by small differences between low values, temperature is first converted to Kelvin and an arbitrary large number (here, 250) is added to daily precipitation values. For temperature, scaling is done by dividing the downscaled value by  $1 + \text{the bias}$  when values fall more than 3% below or above the lowest or highest observed values (in Kelvin), respectively, or more than 2% above the highest observed precipitation value (with 250 added). For precipitation, the downscaling model in some cases predicts values below zero. These are reset to zero.

## 2.6. Variable-specific refinements

Although the downscaling model is purposely designed to be applicable to any variable with a relatively symmetric distribution, predictors must be preselected for each variable and there are some differences in the initial processing of each predictor that can improve the performance of the model in downscaling.

Selection of predictors for temperature and precipitation downscaling has been the subject of several comparative studies (Huth, 1999; Wilby and Wigley, 2000; Widmann *et al.*, 2003; Jeong *et al.*, 2012). ARRMs have been designed to allow for user-selected predictors, if desired. For the purposes of model evaluation and comparison, predictors were chosen to be the same variables as the predictands: 2 m maximum and minimum temperature and 24 h cumulative precipitation. These are the most frequently archived daily output from both CMIP3 and CMIP5 AOGCMs; furthermore, comparison with upper-air predictors for the stations in this study showed no consistent improvement that would affect the performance of

the downscaling model. For models that archive convective, total, and/or large-scale precipitation, the downscaling model calculates the RMSE for the historical training period between the observations and separate downscaled values using each of the three predictors. The predictor variable and corresponding regression model for the training period with the lowest RMSE for a particular month is used to downscale future precipitation for that month and station. This refinement significantly improved the method's ability to simulate precipitation over regions that tend to experience more convective-type precipitation, including the subtropics and mid-latitude summer.

Smoothing AOGCM output has been previously recommended (e.g. Raisanen and Ylhäisi, 2011), and it has been suggested that the smoothing that results from averaging may be one of the reasons why ensemble AOGCM projections typically outperform any individual model simulation (Knutti *et al.*, 2010). Here, temperature fields are smoothed using Empirical Orthogonal Function (EOF) analysis, retaining only the EOFs accounting for 97% of the original variance. Root-mean-square errors (RMSEs) identified 97% as a generally appropriate threshold, with both higher and lower thresholds resulting in higher errors. This step improved model performance, especially for inland stations with higher variance.

Compared to temperature, precipitation tends to display a greater amount of smaller scale variability. This is likely one of the reasons why EOF filtering was found to degrade rather than assist precipitation downscaling. Precipitation is also a combination of a binary (wet/dry) and a continuous non-Gaussian distribution that must be transformed into a more symmetrical distribution before it can be ranked by quantile. Dettinger *et al.* (2004) used the square root of daily precipitation as a predictor, but we found that taking the natural logarithm of precipitation achieves a more symmetric distribution. To address the binary nature of the data, dry days must be omitted from the regression. However, simulated and observed time series of precipitation rarely contain the same amount of precipitable days. To correct for any differences in number of rainy days between observations and the simulated time series, the two time series are ordered by rank, extracting the top number of values in each vector corresponding to the number of rainy days in the shorter non-zero time series (usually observations, because AOGCMs tend to 'drizzle' or simulate many more low-precipitation days than observed; e.g. Chen *et al.*, 1996; Sun *et al.*, 2006; Perkins *et al.*, 2007). Drizzle is also addressed by setting downscaled precipitation amounts less than trace (typically defined as 0.005 inches or 0.127 mm) to zero.

The fact that the downscaling process can only be applied to precipitable days raises concerns regarding model performance in extremely dry regions. Given the typical variance of precipitation, to have a confidence level of 95% there must be at least 57 samples in the dataset (i.e. at least 57 wet days in each of the 12 monthly time series that span the entire training period). This value was determined by applying a simple sample size calculation for linear and logistic regression following

Hsieh *et al.* (1998). During the dry season in arid regions the sample size can be insufficient, even for 50 years of data including half a month on either side. If the sample is insufficient, the model automatically expands it by including an extra week's data on either side of that month (thus containing 3 weeks each of the prior and subsequent months), repeating the process up to a maximum of eight times until a sample size of at least 57 is reached. If 16 weeks have been added and the sample is still less than 57 but greater than 20, a linear regression is used. If less than 20 (which, for a training period of 50 years, would mean less than 1 day in 2 years with measurable precipitation), all downscaled values are set to zero for that month. This procedure has been tested and produces reasonable downscaling of historical precipitation in regions that are arid or semiarid.

The ARRM model was constructed in three distinct phases to quantify the contribution of specific elements to model performance. All phases build monthly models that incorporate 2 weeks' data on either side of the target month to double the sample size, and all versions prefilter the temperature and precipitation predictors as described above before ranking by value. The difference between the versions is the function used to fit the quantile–quantile relationship between observations and historical simulations. The first version applies a least mean squares linear fit (using the function  $lm$  in R), similar to that used in the SAR approach of Dettinger *et al.* (2004). The second version applies the piecewise regression function described above. The third version also uses piecewise regression, but incorporates removal of negative or flat slopes and bias correction near the tails. Removal of negative slopes is not expected to yield significant improvements in model performance, and in some cases it may even degrade initial performance; however, it is necessary to reduce the risk of unrealistic statistical relationships between modelled and observed values. The purpose of the comparison is not primarily to demonstrate the superiority of the final model, but rather to ensure that model performance is not overly degraded by this step.

The three different versions will be referred to as *linear*, *simple piecewise*, and *full piecewise* downscaling models, respectively. The ability of these three versions to downscale daily temperature and precipitation for 20 long-term stations in North America was evaluated using the data and model simulations described next.

### 3. Data and simulations

#### 3.1. Observations

Downscaling was conducted and tested using observed daily minimum and maximum temperature and 24-h cumulative precipitation amounts for 20 long-term North American weather stations for the period 1960–2009. Seventeen of the stations are distributed across diverse climatic regions in the continental United States including coastal, central, and mountainous regions; two stations

are located in Canada; and one in Mexico (Figure 4). Records were obtained from the GHCN (Peterson and Vose, 1997).

Although GHCN station data have already undergone a standardized quality control (Durre *et al.*, 2008), before using the station data for downscaling they were filtered using a quality control algorithm to identify and remove (replacing with 'NA') erroneous values in the GHCN database. This additional quality control step included three tests for errors, removing data on any days where the daily reported minimum temperature exceeds the reported maximum, any temperature values above (below) the highest (lowest) recorded values for North America ( $-50$  to  $70^{\circ}\text{C}$ ) or with precipitation below zero or above the highest recorded value for the continental United States (915 mm in 24 h), and repeated values of more than five consecutive days with identical temperature or non-zero precipitation values to the first decimal. Additionally, an erroneous value was found for Hialeah, FL, of  $-17.8^{\circ}\text{C}$  on 8 November 2003 (with temperatures of  $29.4^{\circ}\text{C}$  the previous day and  $30.0^{\circ}\text{C}$  the following day), which was removed.

### 3.2. Atmosphere–ocean general circulation models

Model output from four AOGCMs was used to evaluate the downscaling model. The models chosen for this study are all part of the Coupled Model Inter-comparison Project version 3 (Meehl *et al.*, 2007): the National Center for Atmospheric Research Community Climate System Model version 3 (CCSM3; Collins *et al.*, 2006), the National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 (GFDL CM2.1; Delworth *et al.*, 2006), the United Kingdom Met Office Climate Model version 3 (HadCM3; Pope *et al.*, 2000), and the Department Of Energy/National Center for Atmospheric Research Parallel Climate Model (PCM; Washington *et al.*, 2000). Previous studies (e.g. Gleckler *et al.*, 2008; Stoner *et al.*, 2009; Rusticucci *et al.*, 2010) show that these models are able to represent key features of atmospheric variability including teleconnection patterns, extreme temperature and precipitation, as well as other climate metrics. A description of the models is provided in Table I.

### 3.3. Scenarios

Historical AOGCM simulations correspond to the CMIP '20th Century Climate in Coupled Models' or 20C3M total forcing scenarios. These scenarios include forcing from anthropogenic emissions of greenhouse gases, aerosols, and reactive species; changes in solar output; particulate emissions from volcanic eruptions; changes in tropospheric and stratospheric ozone; and other influences required to provide a complete picture of the climate over the last century. Where multiple simulations were available, the first was used here (run 1).

To represent a broad range of alternative climate futures, simulations corresponding to the IPCC Special

Report on Emission Scenarios (SRES) higher (A1fi) and lower (B1) emission scenarios were used (Nakićenović *et al.*, 2000). These scenarios describe internally consistent pathways of future societal development and corresponding emissions, with atmospheric  $\text{CO}_2$  concentrations reaching approximately 550 ppm (B1) and 990 ppm (A1fi) by 2100.

20C3M simulations only cover the period 1960–1999, in order to have a longer range of historical simulations we extended this period by 10 years by including 2000–2009 simulated output from the A2 SRES scenario. We find this to be a reasonable approach because the inertia of the climate system delays its response to forcings from increased greenhouse gases and other factors identifying each scenario and there is not much difference between the scenarios over the first decade of the century (Stott and Kettleborough, 2002).

## 4. Model evaluation

### 4.1. Creation of independent simulated historical time series

To evaluate the three versions of ARRM (linear, simple piecewise, and full piecewise), 50 years' worth of data and historical total forcing simulations from 1960 to 2009 were used to build downscaling models for daily temperature and precipitation for 20 long-term weather stations across North America.  $N$ -fold cross-validation, or jackknifing, was used whereby the downscaling model was trained on all but one of the years, then used to predict values for the remaining year. ARRM builds a separate model for each of the 12 months of the year, so this process was repeated until 600 independent 1-month simulated daily time series had been generated for each location, independent of the observations used to train the statistical model. These were then combined into a single 50-year time series for evaluation.

Use of cross-validation in creating the historical simulated time series to be evaluated against observations is a crucial aspect of the evaluation. If the statistical model had been trained on all 50 years and then used to predict those same 50 years, comparing the resulting time series with observations would simply reflect the ability of the regression function to fit the data. The results of such an evaluation would be *improved* by overfitting, for example, by allowing the piecewise regression function to fit an infinite number of knot points to the quantile–quantile relationship. In contrast, by generating an independent time series, the evaluation instead reflects the ability of the model to recreate observations that were not used to train the model. The results of such an evaluation are *degraded* by overfitting that makes the model less generalizable. The split-sample approach, whereby observational data are divided into a training and evaluation period, is commonly used to evaluate statistical downscaling methods in the literature. The ability of the statistical model to reproduce

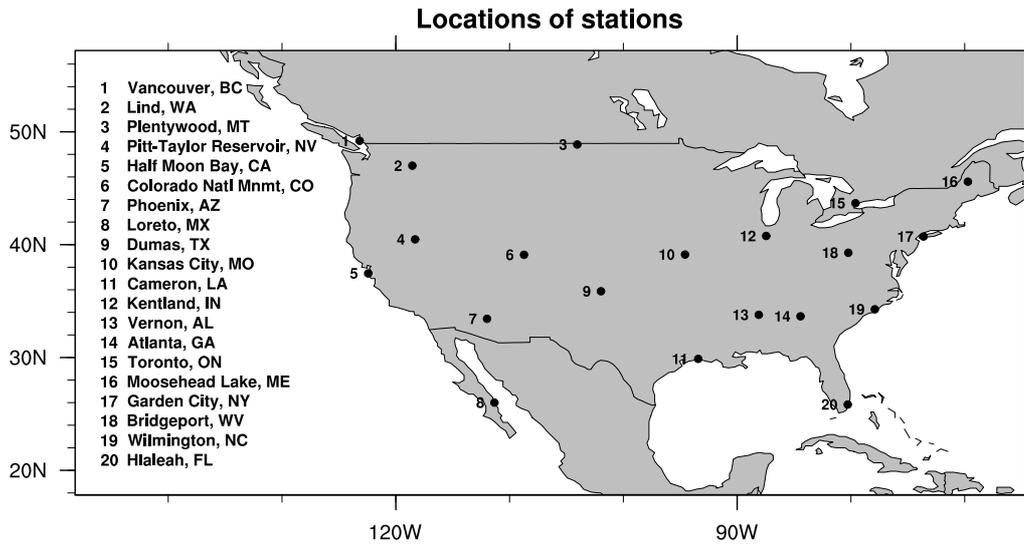


Figure 4. Locations of the 20 North American stations used to evaluate ARRMs.

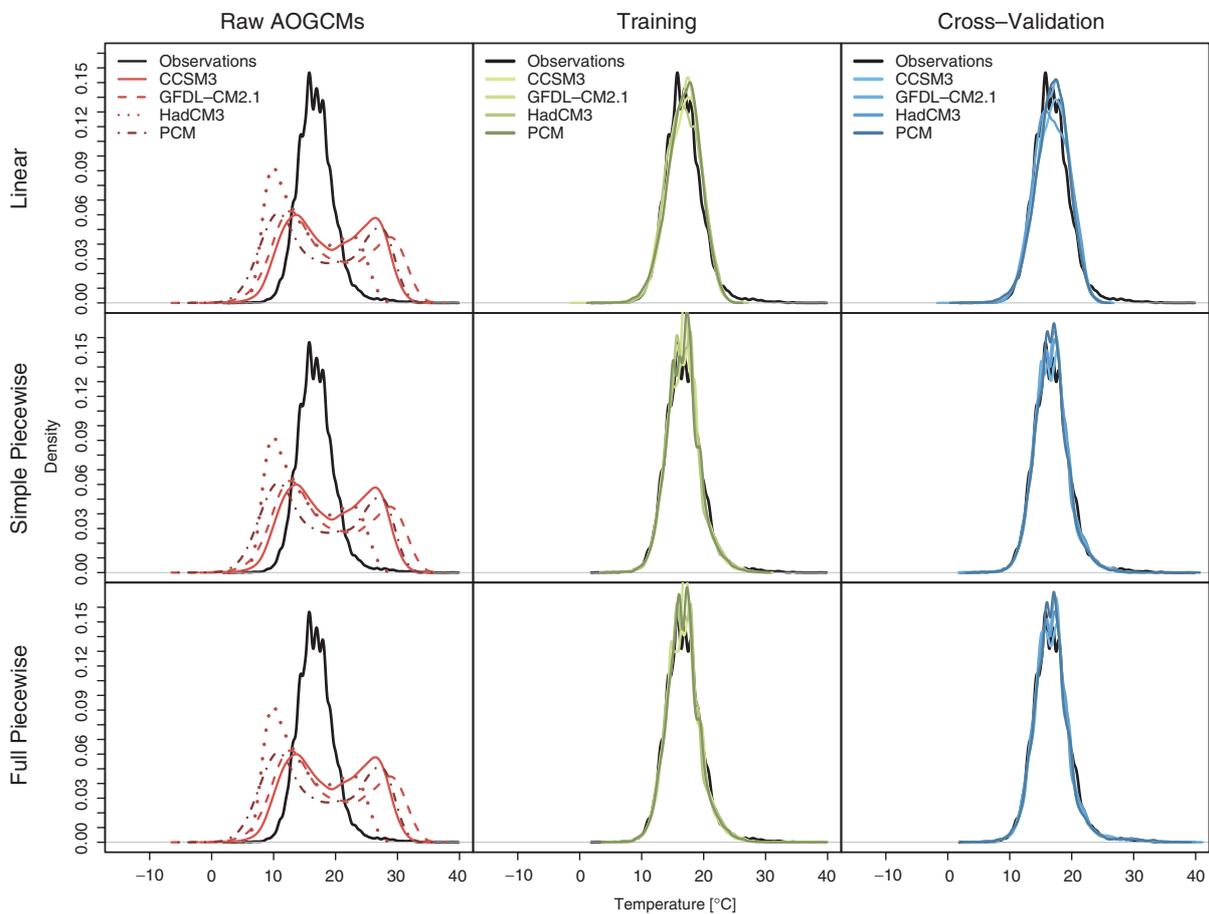


Figure 5. Probability distributions for Half Moon Bay, CA, 1960–2009 observed daily maximum temperature (black) and AOGCM (left column), training (middle column), and cross-validation (right column) simulated daily maximum temperature using linear (top row), simple piecewise (middle row), and full piecewise (bottom row) regression models. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

observed natural variability at a given location, however, depends on the degree to which it is able to sample from that variability in both training and evaluation. The split-sample approach limits the sample size of both the training and observation periods (typically,  $N/2$  years

each), whereas the jackknifed cross-validation approach used here, with a training period of  $N - 1$  years and an evaluation period of  $N$  years, more closely approximates the skill of the full dependent sample model that will be used to downscale future projections. As the purpose

Table I. Summary of key characteristics of AOGCMs used, including acronyms, host institution, as well as atmospheric and oceanic resolution.

Model acronym	Host institution	Atmospheric resolution	Oceanic resolution
CCSM3	NCAR (USA)	1.4° × 1.4°	1.125° × 0.43°
GFDL-CM2.1	NOAA/GFDL (USA)	2.0° × 2.5°	0.9° × 1.0°
HadCM3	UK Met Office (UK)	2.5° × 3.75°	1.25° × 1.25°
PCM	NCAR (USA)	2.81° × 2.81°	1.0° × 1.0°

of downscaling is to ‘recreate’ future conditions that cannot be used to train the model, we argue that the type of evaluation done here is more relevant to assessing the performance of a downscaling model. This is somewhat similar to a bootstrapping approach (Li *et al.*, 2010).

#### 4.2. Evaluating temperature downscaling

The overall skill of the downscaling models is assessed in terms of their ability to reproduce the observed annual distribution (through comparing probability distribution functions), the RMSEs compared to observations, and the absolute value of the bias in the 0.1th, 1st, 10th, 25th, 50th, 75th, 90th, 99th, and 99.9th quantiles. Model projections are also compared (although not evaluated) for end-of-century under the SRES A1fi (higher) and B1 (lower) emissions scenarios.

To gain a qualitative perspective on the downscaling, we first compare observed, AOGCM-simulated (nearest grid cell), downscaled (training period), and downscaled (independent evaluation period) maximum and minimum temperature distributions for the coastal location of Half Moon Bay, CA (Figures 5 and 6), for which the simulated and observed temperature distributions differ noticeably. The three rows correspond to the three versions of the downscaling model (linear, simple piecewise, and full piecewise). The three columns show AOGCM predictions, predictions from training the downscaling model on all 50 years of data, and the independent cross-validation predictions, derived by the method described above. Identical figures for the remaining 19 stations and other graphics not included in this publication are available online (<http://temagami.ttu.edu/arm/>).

For this location, all AOGCMs simulate a wide distribution for maximum temperature with two peaks near 10 and 28 °C (Figure 5). In contrast, the distribution of observed maximum temperatures is narrow and only has one peak near 17 °C. The HadCM3 distribution is additionally skewed towards lower temperatures. One reason for the large difference between observed and simulated distributions is due to the landmask in the AOGCMs, which can have anything from 0 to 100% land fraction in coastal grid cells, differing between AOGCMs. Table II gives land fraction values for grid cells used to downscale stations near the coast. The grid cell downscaled to Half Moon Bay has only partial land coverage in most models (PCM: 15.2%, CCSM3: 53.8%, and GFDL-CM2.1: 84.2%) and is a complete ocean grid cell in the HadCM3 model. Predictions might be improved by selection of a

Table II. Fraction of land (in percent) of AOGCM grid cell used to downscale each station for the four AOGCMs. Values are given only for stations near a coast as the percentage of land in grid cells used to downscale inland stations were all 100%.

Station	CCSM3	GFDL-CM2.1	HadCM3	PCM
Cameron, LA	85	66	100	42
Garden City, NY	91	84	100	44
Half Moon Bay, CA	54	84	0	15
Hialeah, FL	31	10	0	55
Loreto, MX	2	37	0	40
Moosehead Lake, ME	100	95	100	100
Phoenix, AZ	100	100	100	94
Vancouver, BC	100	100	100	88
Wilmington, NC	51	39	100	89

different AOGCM grid cell; however, the purpose here is not to generate optimal predictions but rather to test the ability of the downscaling method to correct AOGCM output. From that perspective, using a near-shore grid cell to simulate coastal conditions represents a greater challenge for the model, and all three versions of the downscaling model are able to approximate observations for these grid cells, narrowing the simulated distribution and removing the double peaks. The linear model is able to capture the general shape of the observed distribution, but underestimates high temperatures towards the tail of the distribution. This is improved upon by the simple piecewise model and almost completely resolved by the full piecewise model. There is little difference between the results for the training (middle column) and independent (last column) predictions, indicating that the downscaling model does not overfit and is successful at simulating observed conditions outside the training period.

There are some differences between maximum and minimum temperature (Figure 6). First, AOGCM distributions for minimum temperature more closely resemble observed, although generally skewed towards cooler when compared with warmer values. Second, all three downscaling models perform well at the tails of the distribution, but the peak of the distribution is better resolved by the two downscaling models that apply the piecewise regression technique.

Figure 7 compares the RMSE in maximum temperature across the entire distribution for all 20 stations. Applying any of the three downscaling models greatly reduces RMSEs compared with raw AOGCM outputs, which in most cases are an order of magnitude larger. Moreover, the downscaling process is able to transform a broad

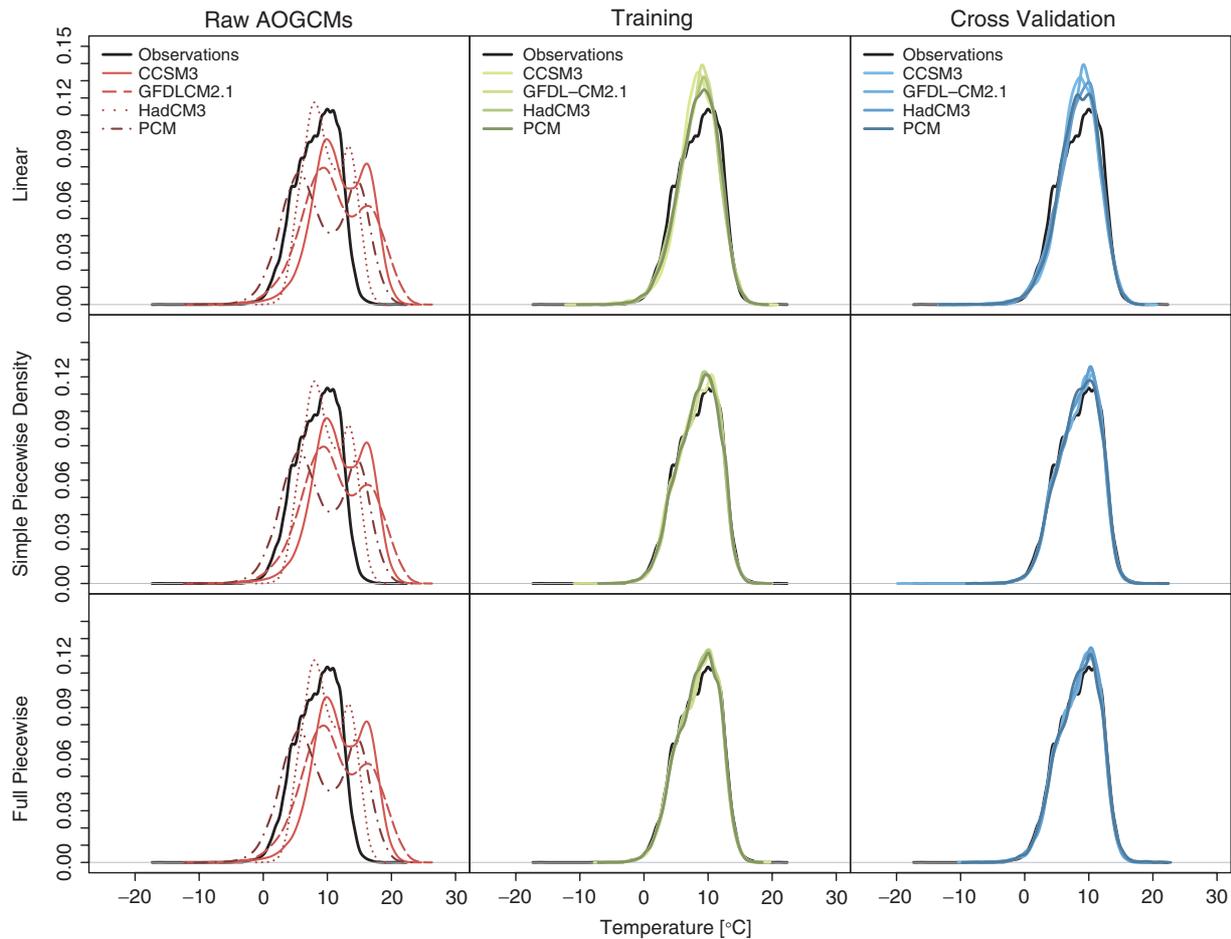


Figure 6. Probability distributions for Half Moon Bay, CA, 1960–2009 observed daily minimum temperature (black) and AOGCM (left column), training (middle column), and cross-validation (right column) simulated daily minimum temperature using linear (top row), simple piecewise (middle row), and full piecewise (bottom row) regression models. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

range of AOGCM predictions into distributions closely resembling observed. For all 20 stations, downscaling reduces the RMSE of simulated historical values from 2 to 8 °C down to less than 0.5 °C. Refining the downscaling technique by applying piecewise regression further decreases the residuals. There is little difference between RMSEs of the simple piecewise and full piecewise regression methods as improvements due to bias removal tend to be offset by removal of negative slopes. Results for minimum temperature (not shown) are similar, except that the RMSE values for AOGCMs tend to be lower, confirming the indication from Figures 5 and 6 that these AOGCMs are generally better at simulating daily minimum when compared with maximum temperatures, regardless of location.

The results of this evaluation are summarized by scatter plots of downscaled versus AOGCM RMSE (Figure 8). Applying downscaling reduces the spread of RMSEs noticeably with the linear version of the downscaling model, and even further when piecewise regression is added to the downscaling model, with RMSE values below 0.5 °C for temperature and below 10 mm for precipitation. For both simple piecewise and full piecewise downscaling models, the majority of points

are clustered between 0.2 and 0.3 °C for temperature and between 1 and 5 mm for precipitation (the far outlier for precipitation for the simple model is HadCM3 downscaled for Phoenix, AZ, with an RMSE value of 76.2 mm), indicating that this level of bias is most likely the limit to the ability of this particular type of statistical downscaling model, within the range of natural variability represented in the training dataset.

The third measure used to evaluate the downscaling methods is by examining the bias in the 0.1th, 1st, 10th, 25th, 50th, 75th, 90th, 99th, and 99.9th quantiles (Figure 9). Bias in AOGCM output is generally an order of magnitude larger than bias downscaled output, regardless of downscaling technique. There is no consistent tendency for AOGCM biases to be larger for certain quantiles, but downscaled quantiles tend to be slightly larger for extreme when compared with median quantiles.

Figure 9 also shows the percentage missing data in the observations for each station. Even for locations with a very high percentage of missing data (Bridgeport and Moosehead Lake) downscaling is able to improve on AOGCM output, although the resulting biases reflect the uncertainty from the very small sample size of the data used to train the statistical models.

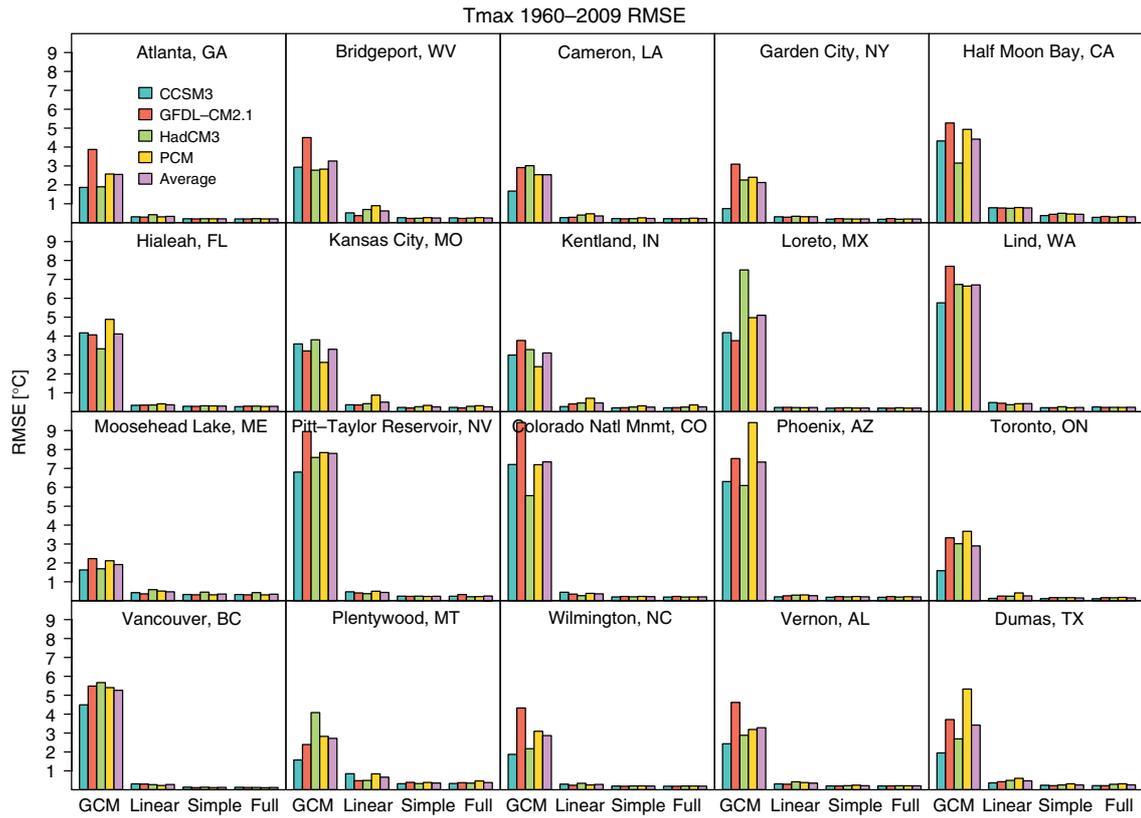


Figure 7. Daily maximum temperature root-mean-square errors, relative to observations, for raw AOGCM simulations as well as linear, simple piecewise, and full piecewise downscaling methods for each of the 20 stations. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

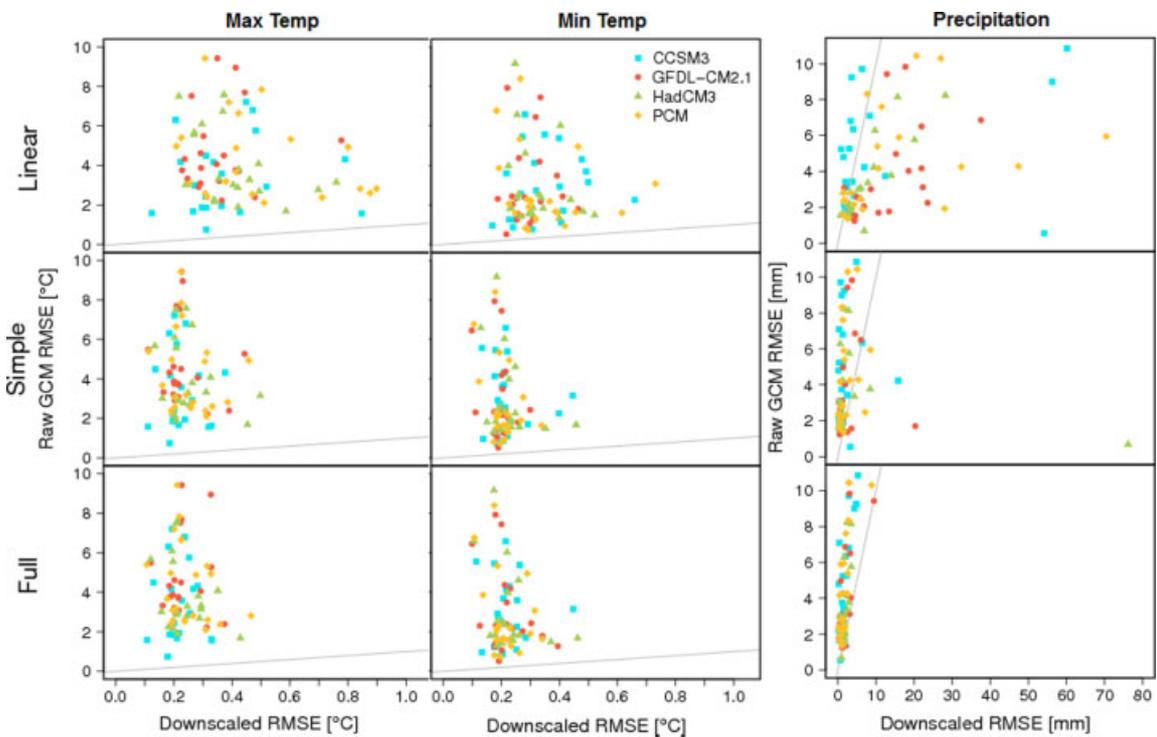


Figure 8. Scatter plot of raw AOGCM 1960–2009 daily maximum, minimum temperature and precipitation RMSE versus cross-validation downscaled daily maximum temperature RMSE for all 20 stations for linear, simple piecewise, and full piecewise downscaling methods. Each dot denotes one station/downscaled AOGCM combination. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

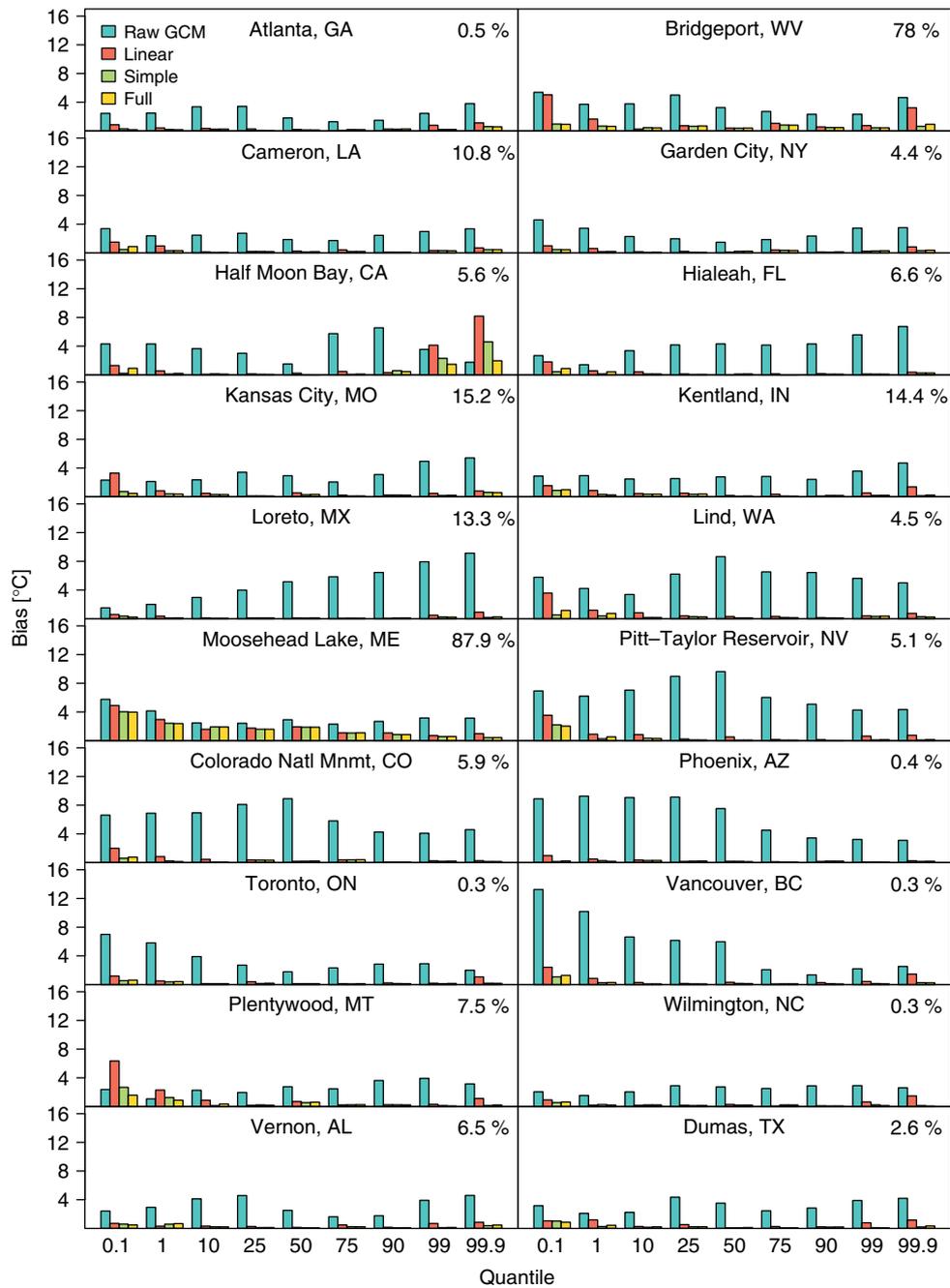


Figure 9. Absolute value of bias in nine quantiles between cross-validation downscaled 1960–2009 daily maximum temperature and observed daily maximum temperature for the same period. Values are averaged across the four AOGCMs and shown for raw, linear, simple piecewise, and full piecewise downscaling methods. The percentage shown in the upper right-hand corner of each plot is the percentage of missing data in the observed record. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

Comparing the reduction in biases in the lowest, middle, and highest quantiles of maximum temperature achieved by downscaling from AOGCM outputs for the cross-validation results shows that using the linear downscaling method noticeably reduces the range in bias relative to AOGCM output for the median quantiles, but not for more extreme quantiles (Figure 10; results for minimum temperature are similar, not shown). Incorporating piecewise regression makes little difference to the 50th quantile when compared with the linear model, but significantly reduces biases in more extreme quantiles.

This suggests that the piecewise regression technique's primary improvement for temperature, compared with a linear model, is in downscaling extreme values.

#### 4.3. Evaluating precipitation downscaling

To gain a qualitative perspective on precipitation downscaling, we first compare observed, AOGCM-based, and downscaled distributions of the natural logarithm of precipitation for 1960–2009 for Kentland, IN (Figure 11). The left column of plots shows the tendency of AOGCMs to drizzle on the left-hand side of the distribution

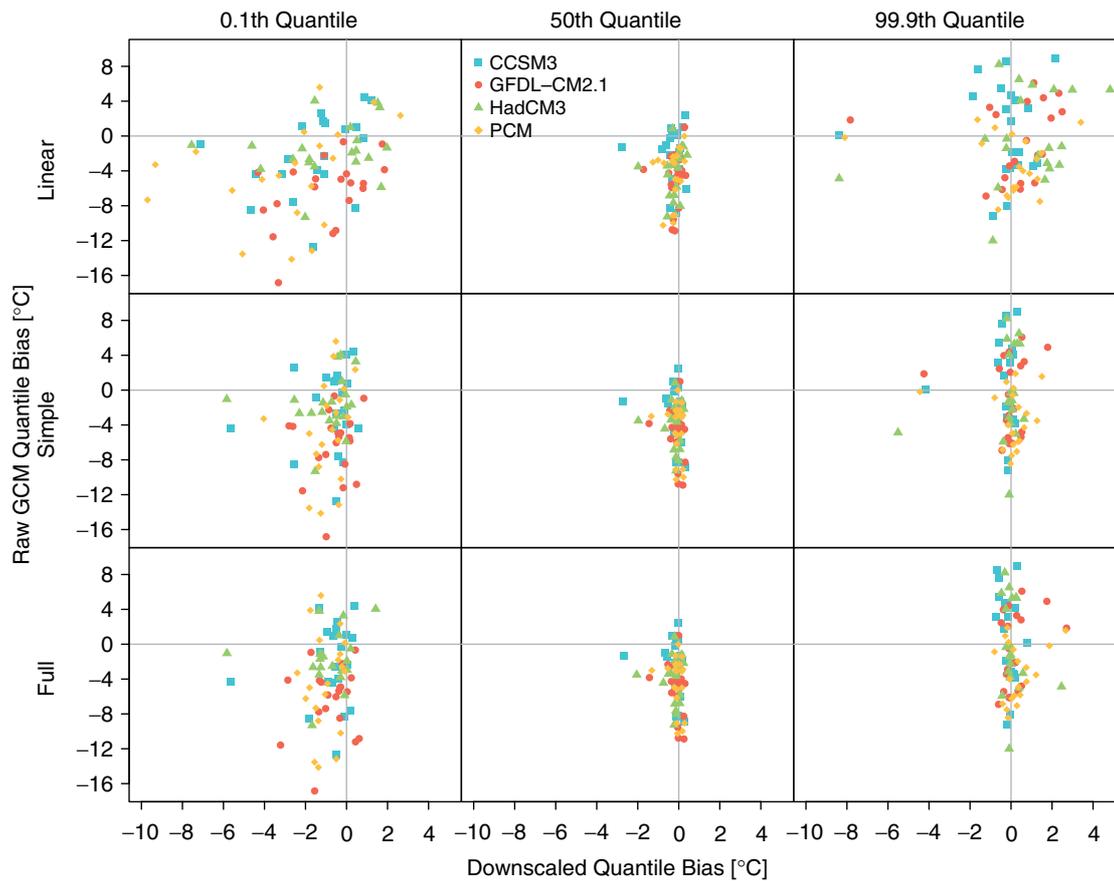


Figure 10. Scatter plots of raw AOGCM maximum temperature bias *versus* downscaled bias for the evaluation period for three separate quantiles representing the tail, the middle, and high ends of the distribution. This relationship is shown for linear (top row), simple piecewise (middle row), and full piecewise (bottom row) downscaling methods. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

and underestimates the magnitude of high-precipitation extremes on the right-hand side of the distribution (e.g. 150 vs 400 mm). The AOGCMs also fail to simulate the double-peaked distribution common to many stations, including Kentland.

The linear version of the downscaling model corrects the lower tail, partly corrects the higher tail (although it introduces some very high-precipitation values), and does not correct for the two peaks in the distribution. Incorporating piecewise regression resolves the peaks but introduces artificially large extreme values that are corrected in the full piecewise method that includes bias correction.

Figure 12 compares RMSE values for all 20 stations between observations, AOGCM output, and downscaled simulations for the evaluation period. For almost all locations, applying the linear downscaling model increases RMSEs relative to AOGCM output. This is most likely due to the linear model simulating extreme values that are too high but carry more weight in the overall RMSE calculations. Piecewise regression corrects the high-end bias and in almost all cases reduces RMSE relative to AOGCM output.

Absolute bias (in percent) in the same nine quantiles as used for temperature (Figure 13) shows that for all nine quantiles, biases are generally small, for the full model

the bars are barely visible for most stations for all nine quantiles. Plotting real-value quantile biases for the 0.1th, 50th, and 99.9th quantiles (Figure 14) shows again that biases are very minimal for the lower and middle quantiles, with larger values for the highest quantile. AOGCM biases in the 99.9th quantile are nearly all negative, i.e. AOGCMs underestimate extreme precipitation accumulation in almost all 20 locations examined here. This is not surprising, given that AOGCM values are averaged over a large area whereas observations are for point sources.

## 5. Future projections

The purpose of most downscaling models is to generate future projections more representative of individual locations than current AOGCMs can provide with grid cell-sized information. Here, we compare the results of AOGCM simulations with ARRM downscaled future projections using the entire historical period (1960–2009) to train each model.

### 5.1. Maximum temperature

Figures 17 and 18 show the change in downscaled *versus* raw AOGCM daily maximum temperature for 2070–2099 relative to the historical period observations

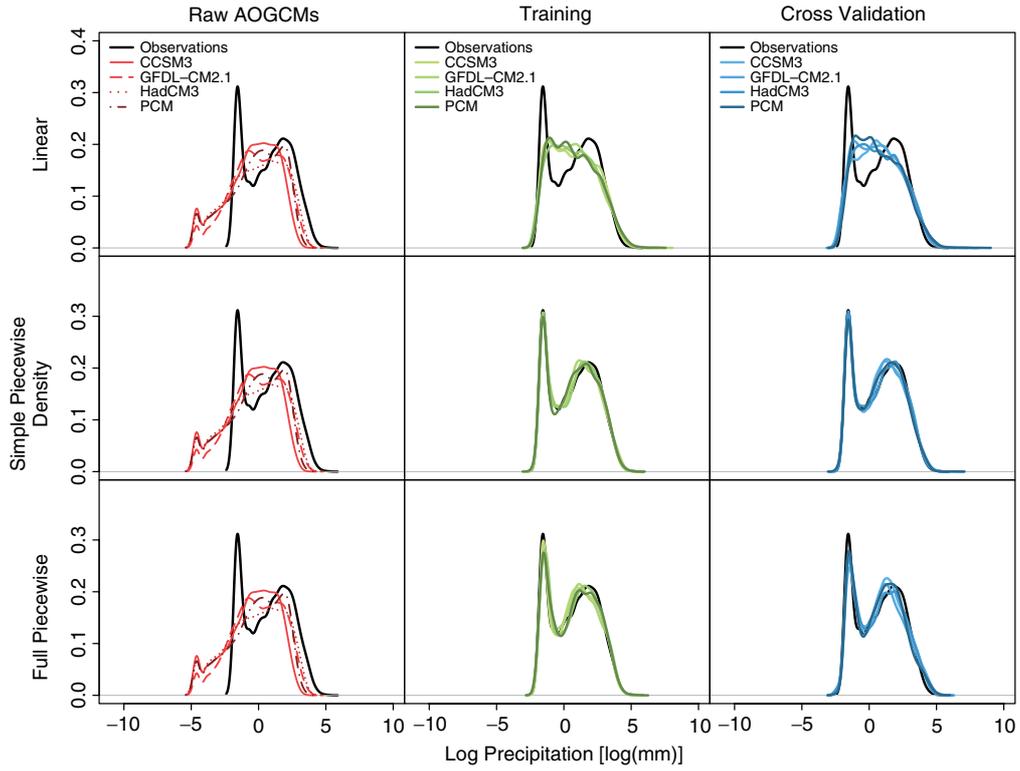


Figure 11. Probability distributions for 1960–2009 observed daily precipitation accumulation (black lines), AOGCM simulated daily precipitation accumulation (left column), training (middle column), and cross-validation (right column) for Kentland, IN. Downscaled simulations are shown for linear (top row), simple piecewise (middle row), and full piecewise (bottom row) methods. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

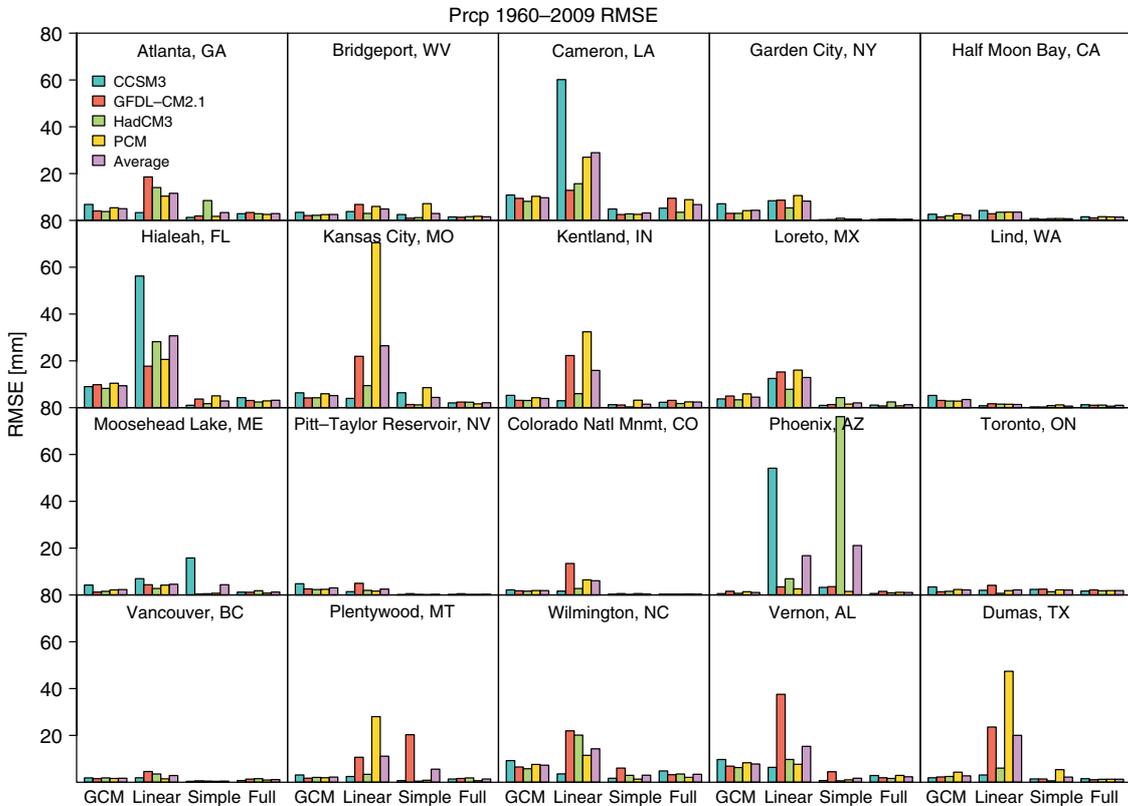


Figure 12. Daily precipitation root-mean-square errors, relative to observations, for raw AOGCM simulations as well as linear, simple piecewise, and full piecewise downscaling methods for the evaluation period for each of the 20 stations. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

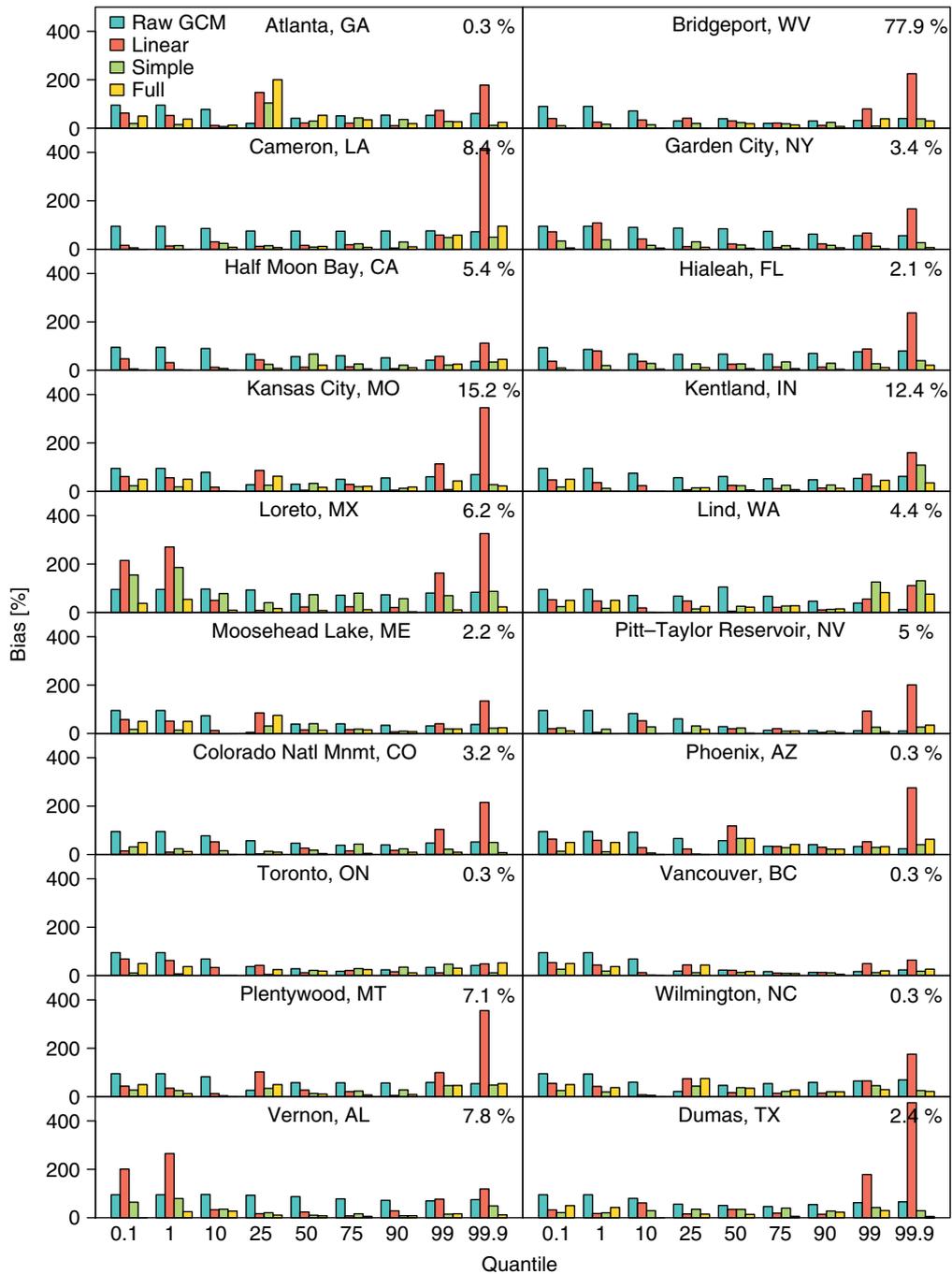


Figure 13. Percentage bias in nine quantiles between cross-validation downscaled and observed 1960–2009 daily precipitation accumulation. Values are averaged across the four AOGCMs and shown for raw, linear, simple piecewise, and full piecewise downscaling methods. The percentage shown in the upper right-hand corner of each plot is the percentage of missing data in the observed record. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

(1960–2009) for the three temperature downscaling models (linear, simple piecewise, and full piecewise) and 0.1th, 50th, and 99.9th quantiles. Under the higher A1fi scenario (Figure 15), the most obvious difference between raw AOGCM *versus* downscaled future changes is that downscaling produces only positive changes (i.e. increases) in all three quantiles illustrated [with the exception of one station (Half Moon Bay, CA) for the linear model and 99.9th quantile], whereas raw AOGCM changes are both positive as well as negative for these

three quantiles, indicating that the raw output projects warming for some locations and cooling for others at the end of the century. For the lower B1 scenario (Figure 16), more stations also show warming at the end of the century after downscaling compared with raw AOGCM results, especially for the middle and upper quantiles. Some cooling is projected for the lowest quantiles, indicating that some stations might see a wider distribution of daily maximum temperature at the end of the century with more extremes in both ends of the distribution.

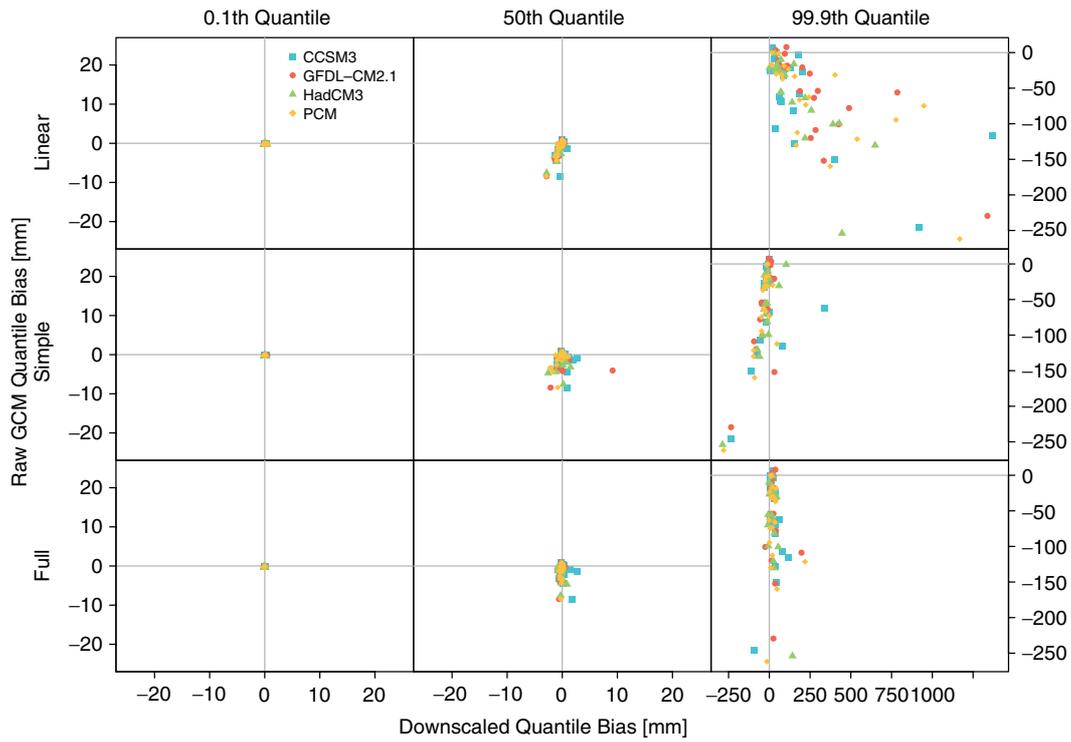


Figure 14. Scatter plots of raw AOGCM quantile bias *versus* downscaled quantile bias for three separate quantiles representing the low tail, the middle, and high tail of the distribution. This relationship is shown for linear, simple piecewise, and full piecewise downsampling methods for the evaluation period. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

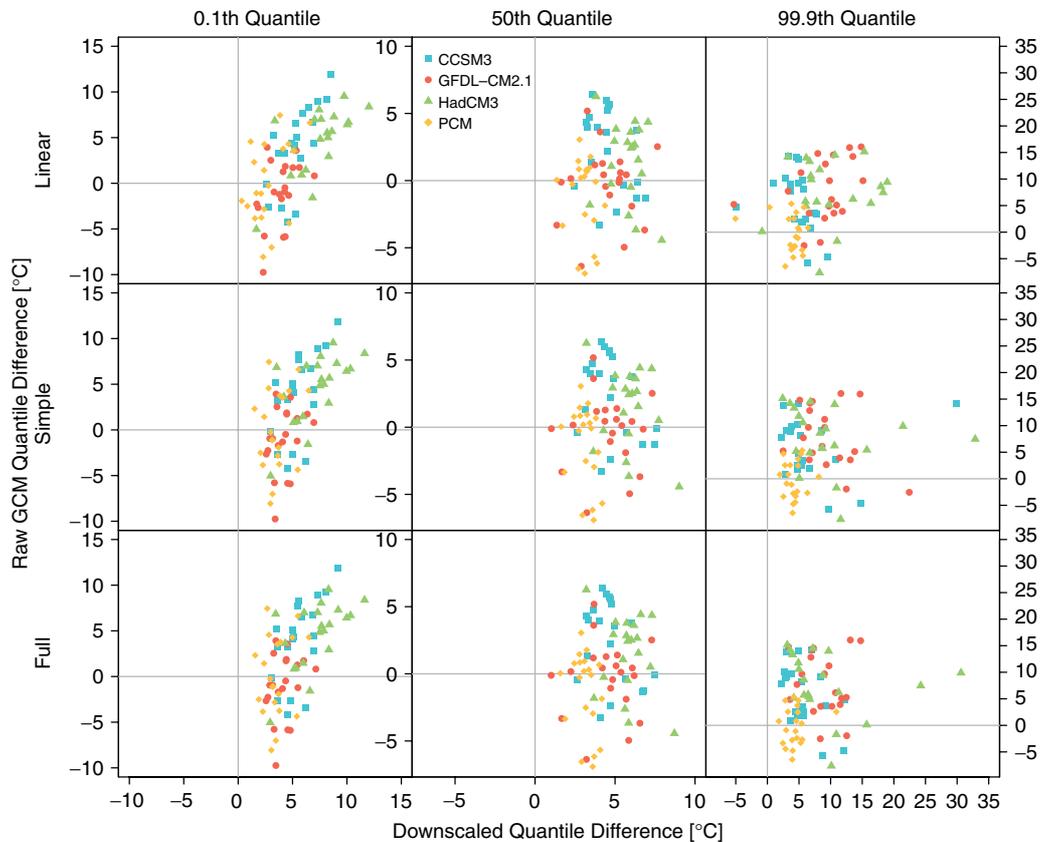


Figure 15. Scatter plots of raw A1fi AOGCM maximum temperature (2070–2099)–(1960–2009) change *versus* downscaled change for three separate quantiles representing the tail, the middle, and high ends of the distribution. This relationship is shown for linear (top), simple piecewise (middle), and full piecewise (bottom) downsampling methods. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

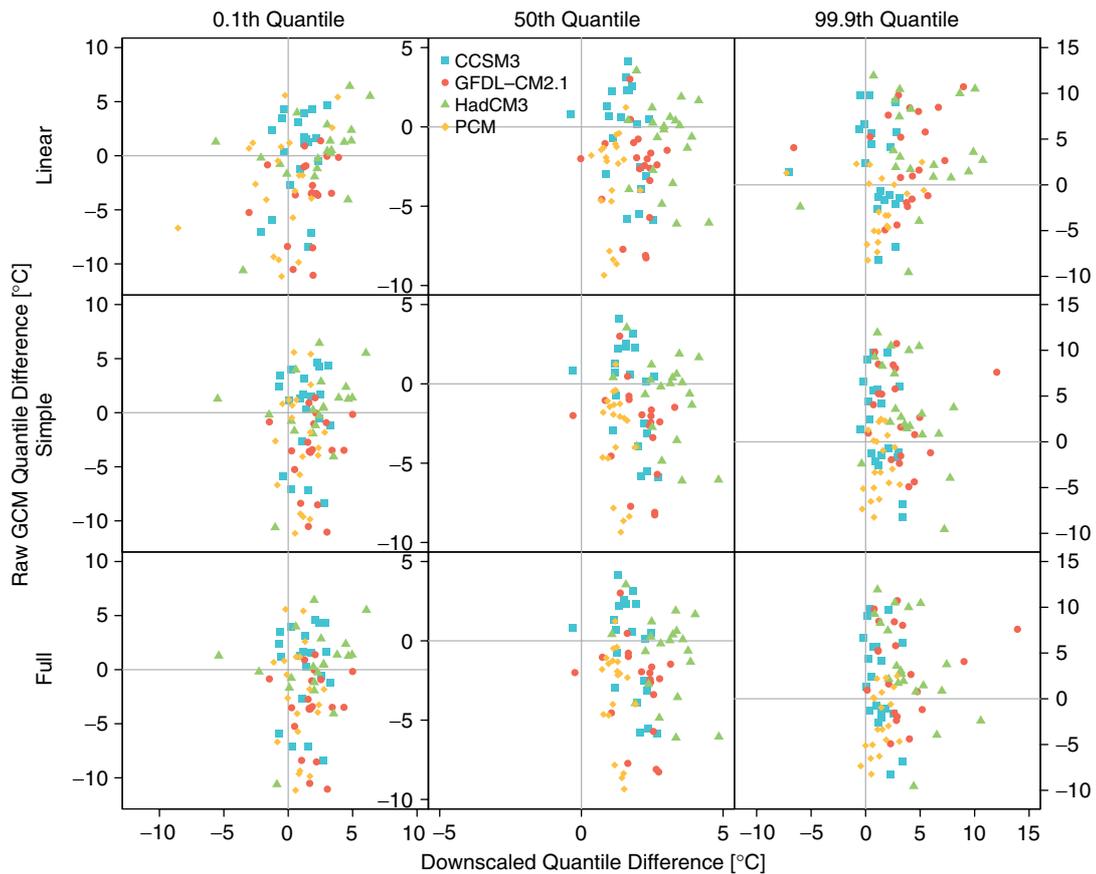


Figure 16. Scatter plots of raw B1 AOGCM maximum temperature (2070–2099)–(1960–2009) change *versus* downscaled change for three separate quantiles representing the tail, the middle, and high ends of the distribution. This relationship is shown for linear (top), simple piecewise (middle), and full piecewise (bottom) downscaling methods. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

Figure 17 shows the mean AOGCM absolute 2070–2099 daily maximum temperature changes, relative to 1960–2009, in each of the nine quantiles for the A1fi and B1 scenarios for raw (light-coloured bars) and downscaled output (dark-coloured bars). There is overall a general agreement among the 20 stations that a greater change in the 50th quantile is projected for the A1fi scenario than for the B1 scenario (Loreto, MX, and Hialeah, FL, being the only exceptions – note that these stations also have low land fraction in all four AOGCMs). However, there is no general tendency for the mean change to be more or less for downscaled *versus* raw AOGCM output, with some locations, such as Atlanta, GA, and Bridgeport, WV, showing a larger change projected for the A1fi scenario than for the B1 scenario (Loreto, MX, and Hialeah, FL, being the only exceptions). Similarly, projected changes in higher quantiles from raw AOGCM can be higher than downscaled for certain locations and lower for others. This indicates that downscaling produces results specific to each location, as opposed to the more general AOGCM grid cell output.

## 5.2. Precipitation

Figure 18 shows the 2070–2099 relative to 1960–2009 raw AOGCM *versus* downscaled precipitation changes in the 0.1th, 50th, and 99.9th quantiles for the three versions

of the downscaling model for the A1fi scenario, given as RMSE differences. Unsurprisingly, there is less than 1 mm change in predicted changes for the 0.1th quantile for all 20 stations for both raw AOGCM and downscaled projections. The reason for the fixed RMSE values for the 0.1th quantile for the raw AOGCMs is due to weather stations not reporting *trace* precipitation, which is set at 0.005 inches (0.127 mm). The higher frequency of low precipitation events, compared with higher precipitation events, in most locations causes the 0.1th quantile to almost always equal to the lowest recorded or simulated precipitation value. The lowest simulated value in AOGCMs, when rounded to the nearest 2 decimals, is 0.01 mm, because AOGCMs do not allow for ‘trace’. All but one stations have a lowest value of 0.2 mm (when converted from inches), whereas one station (Bridgeport, WV) has 0.1 mm as the lowest recorded value, which is the cause for that station not being in agreement with the others in the bias plot (Figure 18). Under both scenarios, AOGCM outputs project little to some (up to about 8 mm) decrease in the amount of precipitation comprising the median quantile, whereas when downscaled the same quantile shows less than 2 mm change from current conditions, with few exceptions, for both scenarios. The largest change is in the 99.9th quantile for both scenarios, with up to several hundred millimetres change

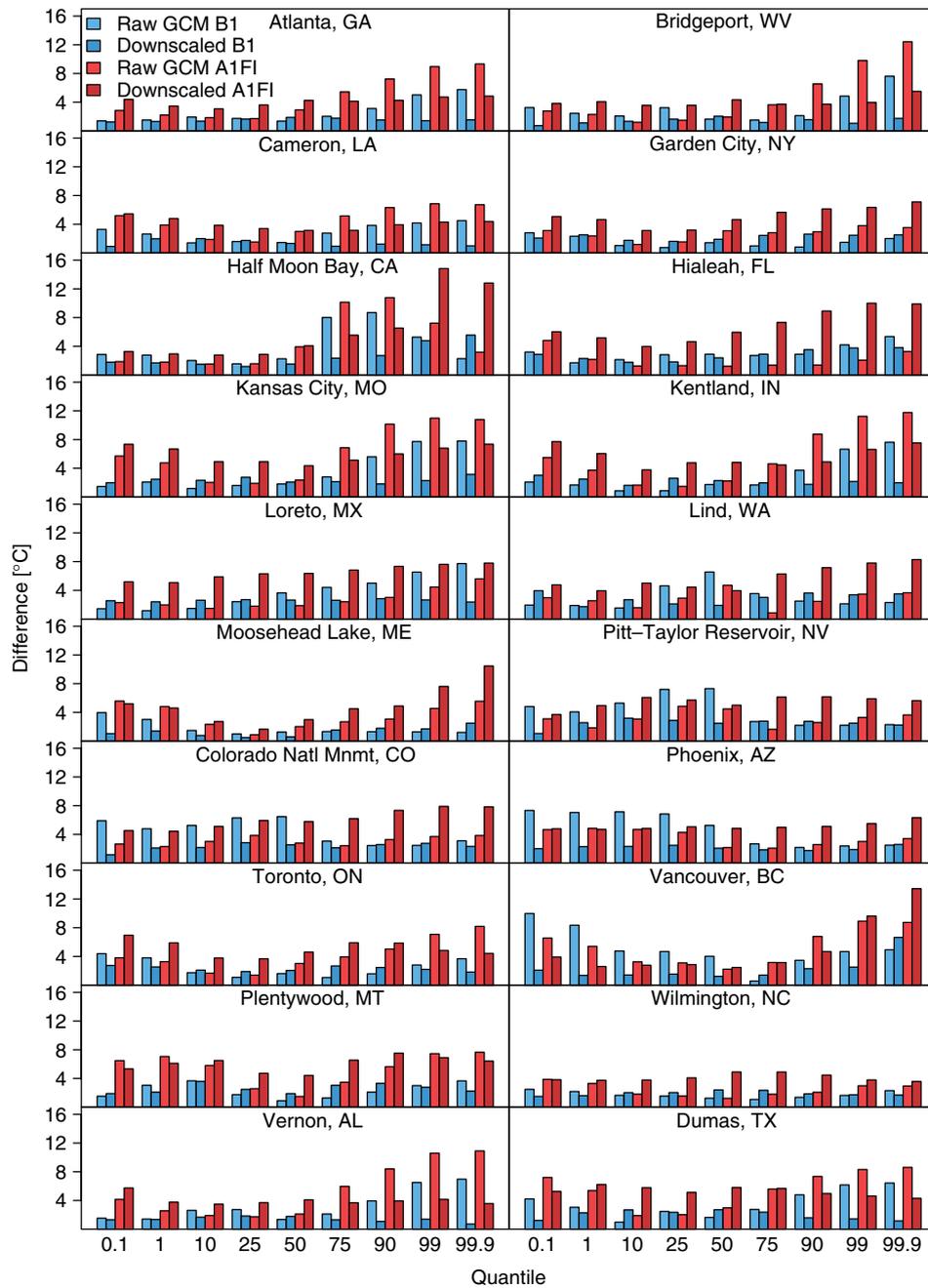


Figure 17. Absolute change in nine quantiles between 2070–2099 simulated and 1960–2009 observed daily maximum temperature. Values are averaged across the four AOGCMs and shown for raw and downscaled (full piecewise) A1fi and B1 scenarios. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

from current extreme conditions. For raw AOGCM projections, future extreme precipitation amounts appear to have decreased, whereas when downscaled, the same locations show a large increase in precipitation extremes, especially for the A1fi scenario (Figure 18). This is most likely due to poor simulation of precipitation at the local scale by AOGCMs and is corrected by applying the statistical downscaling model, which is trained on historical temporal precipitation variability for each location. The linear version of the downscaling model produces very large, up to about 1750 mm, increases in extreme precipitation events, whereas the full piecewise

downscaling model produces more moderate, but still large – up to about 300 mm increases in extreme events. The numbers are very similar, although slightly smaller, for the B1 scenario (not shown here, but available at <http://temagami.ttu.edu/arm/>).

Absolute precipitation changes for the 99th and 99.9th quantiles are shown in Figure 19 for both A1FI and B1 scenarios, averaged across all four AOGCMs. For some stations, such as Hialeah, FL, Loreto, MX, and Vernon, AL, there is a substantial difference between extreme event projections for raw *versus* downscaled AOGCMs, with the raw AOGCM generally projecting

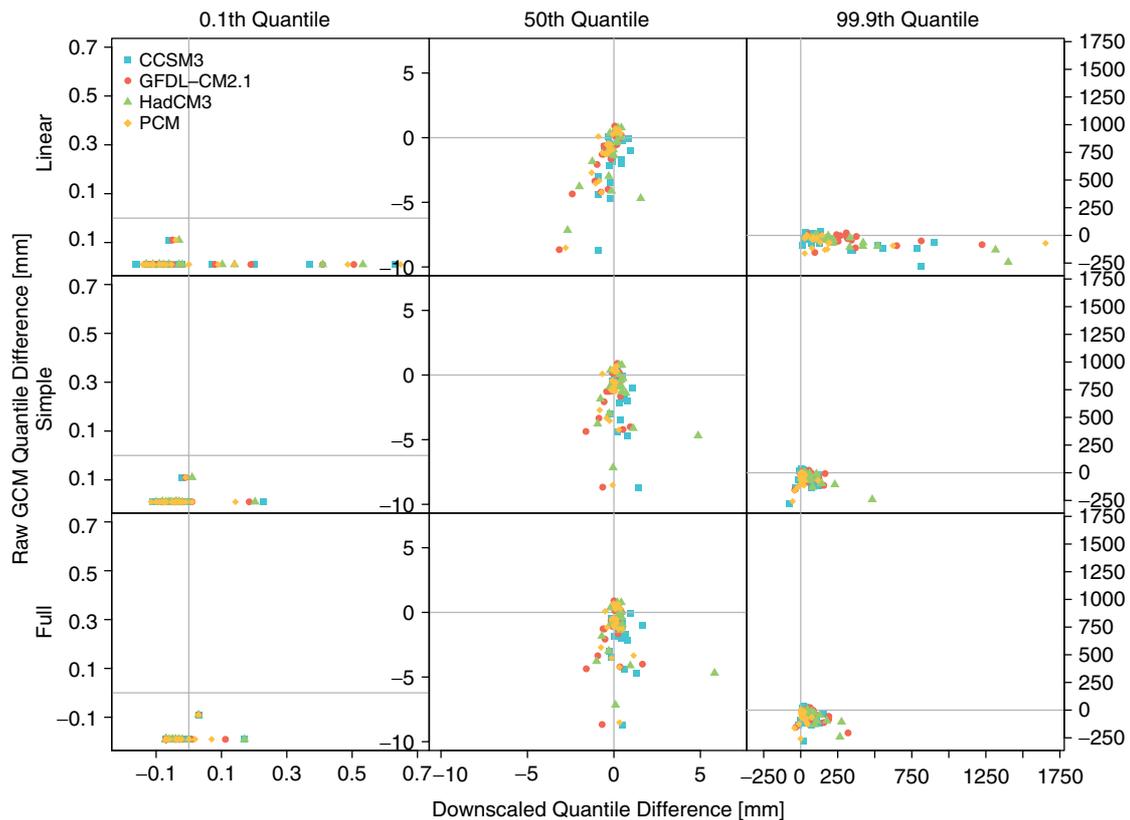


Figure 18. Scatter plots of raw AOGCM A1fi quantile difference between 2070–2099 simulated and 1960–2009 observed daily precipitation *versus* downscaled quantile difference for three separate quantiles representing the low tail, the middle, and high tail of the distribution. This relationship is shown for linear, simple piecewise, and full piecewise downscaling methods. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

larger absolute changes, compared with present conditions, than downscaled projections. The results of the cross-validation evaluation suggest that more confidence could be placed in the downscaled projections when compared with raw AOGCM output, because downscaled projections are tailored to each individual location.

## 6. Conclusions

The ARRM is an empirical statistical downscaling model capable of downscaling local projections of temperature and precipitation to both station-based observations and spatially gridded observations. Quantile regression, the method on which ARRM is based, is unique in that it builds a regression model based on matching the quantiles of the observed and simulated time series as opposed to matching corresponding day-to-day data points, which is the basis for many other regression-based statistical downscaling studies (Wilby *et al.*, 1998; Huth, 1999, 2002; Wilby and Wigley, 2000; Huth *et al.*, 2001; Boé *et al.*, 2007; Kostopoulou *et al.*, 2007). ARRM adds to this by using a piecewise regression model instead of a straight linear regression, which improves its ability to simulate more extreme temperatures and precipitation, one of the major issues with other downscaling methods (Huth, 1999; Goodess *et al.*, 2012).

The downscaling model was evaluated based on cross-validation of three different (linear, simple piecewise, and full piecewise) versions of both the temperature and precipitation models. Each version was evaluated in terms of three different quantities: the distributions, giving a visual image of the skill each model; the RMSE; and bias in a range of quantiles.

The addition of piecewise regression, instead of straight linear regression, was found to have the largest impact on the performance of the method. The largest biases were found to be near the tails of the distribution, primarily due to data sparseness. Some sensitivity to station location was found in the linear versions of the downscaling model, but the addition of piecewise regression was able to eliminate much of this.

For future projections, the spread among projected temperature increases is generally narrower for downscaled temperature compared with raw AOGCM projections for the three quantiles shown, with more stations showing positive temperature changes after downscaling than before, for both higher A1fi and lower B1 scenarios. Downscaled projections of precipitation show smaller changes for the 50th quantile than raw AOGCM projections, for both A1fi and B1 scenarios, but slightly larger changes in extreme events, with projected changes being generally greater under the A1fi scenario than the B1 scenario.

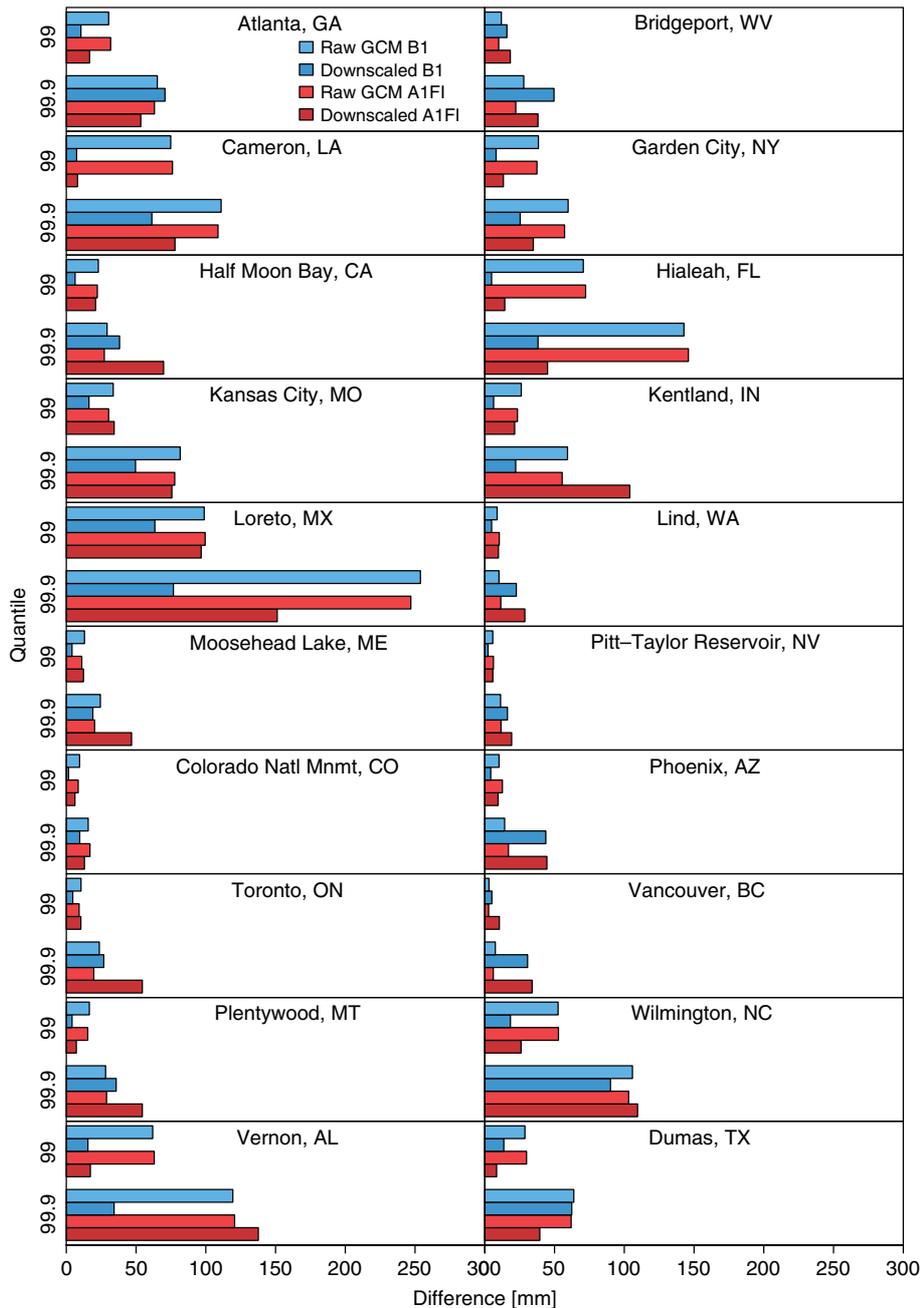


Figure 19. Absolute bias in the 99th and 99.9th quantiles between 2070–2099 simulated and 1960–2009 observed daily precipitation accumulation. Values are averaged across the four AOGCMs. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

Evaluating the ability of ARRM to reproduce observed temperature and precipitation at 20 stations across North America shows that the statistical downscaling model is able to reproduce values from the 0.1th to the 99.9th quantiles with biases generally below 1 °C and 5 mm. Downscaling future projections can alter the sign of AOGCM-simulated changes and usually narrows the range of projected changes across multiple AOGCM simulations.

The ultimate purpose of the ARRM framework is to allow for user selection from a broad range of predictors and predictands to efficiently downscale either point source or gridded observations of any observed

climate variable with a Gaussian-like distribution that can be predicted from large-scale AOGCM output fields. Model performance for station-based temperature and precipitation downscaling appears sufficient to support continued development of such a generalized model. Future work will describe the application of this model to gridded datasets and to downscaling solar radiation and relative humidity.

**Acknowledgements**

This work was supported by the National Science Foundation, under NSF contract number DMS-0724377,

and the U.S. Geological Survey, under contract number G10AC00248. Station data were obtained from the Global Historical Climatology Network, and AOGCM output fields were retrieved from the Earth System Grid.

## References

- Boé J, Terray L, Habets F, Martin E. 2007. Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology* **27**: 1643–1655. DOI: 10.1002/joc.1602.
- Chen M, Dickinson RE, Zeng X, Hahmann AN. 1996. Comparison of precipitation observed over the continental United States to that simulated by a climate model. *Journal of Climate* **9**: 2233–2249. DOI: 10.1175/1520-0442(1996)009<2233:COPOOT>2.0.CO;2.
- Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, McKenna DS, Santer BD, Smith RD. 2006. The Community Climate System Model Version 3 (CCSM3). *Journal of Climate* **19**: 2122–2143. DOI: 10.1175/JCLI3761.1.
- Crane RG, Hewitson BC. 1998. Doubled CO<sub>2</sub> precipitation changes for the Susquehanna basin: down-scaling from the GENESIS general circulation model. *International Journal of Climatology* **18**: 65–76. DOI: 10.1002/(SICI)1097-0088(199801)18:1<65::AID-JOC222>3.0.CO;2-9.
- Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW, Findell KL, Ginoux P, Gnanadesikan A, Gordon CT, Griffies SM, Gudgel R, Harrison MJ, Held IM, Hemler RS, Horowitz LW, Klein SA, Knutson TR, Kushner PJ, Langenhorst AR, Lee HC, Lin SJ, Lu J, Malyshev SL, Milly PCD, Ramaswamy V, Russell J, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Wyman B, Zeng F, Zhang R. 2006. GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate – Special Section* **19**: 643–674. DOI: 10.1175/JCLI3629.1.
- Dettinger MD, Cayan DR, Meyer MK, Jeton AE. 2004. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American river basins, Sierra Nevada, California, 1900–2099. *Climatic Change* **62**: 283–317.
- Durre I, Vose RS, Wuertz DB. 2008. Robust automated quality assurance of radiosonde temperatures. *Journal of Applied Meteorology and Climatology* **47**: 2081–2095.
- Gleckler PJ, Taylor KE, Doutriaux C. 2008. Performance metrics for climate models. *Journal of Geophysical Research* **113**: D06104. DOI: 10.1029/2007JD008972.
- Goodess CM, Anagnostopoulou C, Bárdossy A, Frei C, Harpham C, Haylock MR, Hundecha Y, Maheras P, Ribalaygua J, Schmidli J, Schmith T, Tolika K, Tomozeiu R, Wilby RL. 2012. An intercomparison of statistical downscaling methods for Europe and European regions – assessing their performance with respect to extreme temperature and precipitation events. University of East Anglia: UK, Climate Research Unit Research Publications. Report No. CRU RP11:1–72.
- Hay LE, Wilby RL, Leavesley GH. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *Journal of the American Water Resources Association* **36**: 387–397. DOI: 10.1111/j.1752-1688.2000.tb04276.x.
- Hayhoe K, Cayan D, Field CB, Frumhoff PC, Maurer EP, Miller NL, Moser SC, Schneider SH, Nicholas Cahill K, Cleland EE, Dale L, Drapak R, Hanemann RM, Kalkstein LS, Lenihan J, Lunch CK, Neilson RP, Sheridan SC, Verville JH. 2004. Emissions pathways, climate change, and impacts on California. *Proceedings of the National Academy of Sciences* **101**: 12422–12427.
- Hayhoe K, Wake C, Anderson B, Liang XZ, Maurer E, Zhu J, Bradbury J, DeGaetano A, Stoner AM, Wuebbles D. 2008. Regional climate change projections for the Northeast USA. *Mitigation and Adaptation Strategies for Global Change* **13**: 425–436. DOI: 10.1007/s11027-007-9133-2.
- Hayhoe K, VanDorn J, Croley TC II, Schlegal N, Wuebbles D. 2010. Regional climate change projections for Chicago and the US Great Lakes. *Journal of Great Lakes Research* **36**: 7–21. DOI: 10.1016/j.jglr.2010.03.012.
- Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM. 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* **26**: 1397–1415. DOI: 10.1002/joc.1318.
- Hidalgo HG, Dettinger MD, Cayan DR. 2008. Downscaling with constructed analogues: daily precipitation and temperature fields over the United States. California Energy Commission: California. Report No. CEC-500-2007-123: 1–62.
- Hsieh FY, Bloch DA, Larsen MD. 1998. A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine* **17**: 1623–1634.
- Huth R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research* **13**: 91–101.
- Huth R. 2002. Statistical downscaling of daily temperature in central Europe. *Journal of Climate* **15**: 1732–1742. DOI: 10.1175/1520-0442(2002)015<1731:SDODTI>2.0.CO;2.
- Huth R, Kysely J, Dubrovsky M. 2001. Time structure of observed, GCM-simulated, downscaled and stochastically generated daily temperature series. *Journal of Climate* **6**: 4047–4061. DOI: 10.1175/1520-0442(2001)014<4047:TSSOOGS>2.0.CO;2.
- Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P. 2012. CGCM3 predictors used for daily temperature and precipitation downscaling in Southern Quebec, Canada. *Theoretical and Applied Climatology* **107**: 389–406. DOI: 10.1007/s00704-011-0490-0.
- Knutti R, Furrer R, Tebaldi C, Cernak J, Meehl GA. 2010. Challenges in combining projections from multiple climate models. *Journal of Climate* **23**: 2739–2758. DOI: 10.1175/2009JCLI3361.1.
- Koenker R, Bassett G Jr. 1978. Regression quantiles. *Econometrica* **46**: 33–50.
- Koenker R, Hallock KF. 2001. Quantile regression. *Journal of Economic Perspectives* **15**: 143–156.
- Kostopoulou E, Giannakopoulos C, Anagnostopoulou C, Tolika K, Maheras P, Vafiadis M, Founda D. 2007. Simulating maximum and minimum temperature over Greece: a comparison of three downscaling techniques. *Theoretical and Applied Climatology* **90**: 65–82. DOI: 10.1007/s00704-006-0269-x.
- Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research* **115**: D10101. DOI: 10.1029/2009JD012882.
- Luers AL, Cayan DR, Franco G, Hanemann M, Croes B. 2006. Our changing climate, assessing the risks to California. California Energy Commission: California. Report no. CEC-500-2006-077: 1–16.
- Mearns LO, Gutowski WJ, Jones R, Leung LY, McGinnis S, Nunes AMB, Qian Y. 2009. A regional climate change assessment program for North America. *Eos* **90**: 311–312. DOI: 10.1029/2009EO360002.
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE. 2007. The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bulletin of the American Meteorological Society* **88**: 1383–1394. DOI: 10.1175/BAMS-88-9-1383.
- Nakićenović N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S, Gregory K, Grübler A, Jung TY, Kram T, La Rovere EL, Michaelis L, Mori S, Morita T, Pepper W, Pitcher H, Price L, Riahi K, Roehrl A, Rogner HH, Sankovski A, Schlesinger M, Shukla P, Smith S, Swart R, van Rooijen S, Victor N, Dadi Z. 2000. Special Report on Emission Scenarios. *Intergovernmental Panel on Climate Change*, Retrieved 1 April 2011. Available at [http://www.grida.no/publications/other/ipcc\\_sr/](http://www.grida.no/publications/other/ipcc_sr/)
- O'Brien TP, Sornette D, McPherron RL. 2001. Statistical asynchronous regression: determining the relationship between two quantities that are not measured simultaneously. *Journal of Geophysical Research* **106**: 13247–13259. DOI: 10.1029/2000JA900193.
- Perkins SE, Pitman AJ, Holbrook NJ, McAneney J. 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature and precipitation over Australia using probability density functions. *Journal of Climate* **20**: 4356–4376. DOI: 10.1175/JCLI4253.1.
- Peterson TC, Vose RS. 1997. An overview of the Global Historical Climatology Network temperature database. *Bulletin of the American Meteorological Society* **78**: 2837–2849. DOI: 10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2.
- Pope VD, Gallani ML, Rowntree PR, Stratton RA. 2000. The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics* **16**: 123–146. DOI: 10.1007/s003820050009.
- R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. Available at <http://www.R-project.org/>

- Raisanen J, Ylhaisi JS. 2011. How much should climate model output be smoothed in space? *Journal of Climate* **24**: 867–880. DOI: 10.1175/2010JCLI3872.1.
- Rusticucci M, Marengo J, Penalba O, Renom M. 2010. An intercomparison of model-simulated in extreme rainfall and temperature events during the last half of the twentieth century. Part 1: Mean values and variability. *Climatic Change* **98**: 493–508. DOI: 10.1007/s10584-009-9742-8.
- Schmidli J, Frei C, Vidale PL. 2006. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology* **26**: 679–689. DOI: 10.1002/joc.1287.
- Stehlik J, Bardossy A. 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *Journal of Hydrology* **256**: 120–141. DOI: 10.1016/S0022-1694(01)00529-7.
- Stoner AMK, Hayhoe K, Wuebbles DJ. 2009. Assessing general circulation model simulations of atmospheric teleconnection patterns. *Journal of Climate* **22**: 4348–4372. DOI: 10.1175/2009JCLI2577.1.
- Stott PA, Kettleborough JA. 2002. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**: 723–726. DOI: 10.1038/416723a.
- Sun Y, Solomon S, Dai A, Portman RW. 2006. How often does it rain? *Journal of Climate* **19**: 916–934. DOI: 10.1175/JCLI3672.1.
- USGCRP: National Assessment Synthesis Team. 2000. Climate change impacts on the United States: the potential consequences of climate variability and change. Cambridge University Press: Cambridge, UK. 620.
- USGCRP: Karl TR, Melillo JM, Peterson TC (eds). 2009. Global Climate Change Impacts in the United States. Cambridge University Press: Cambridge, UK; 196.
- Washington WM, Weatherly JW, Meehl GA, Semtner AJ, Bettge TW, Craig AP, Strand WG, Arblaster J, Wayland VB, James R, Zhang Y. 2000. Parallel climate model (PCM) control and transient simulations. *Climate Dynamics* **16**: 755–774. DOI: 10.1007/s003820000079.
- Widmann M, Bretherton CS, Salathe EP Jr. 2003. Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor. *Journal of Climate* **16**: 799–816. DOI: 10.1175/1520-0442(2003)016<0799:SPDOTN>2.0.CO;2.
- Wilby RL, Wigley TML. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* **20**: 641–661. DOI: 10.1002/(SICI)1097-0088(200005)20:6<641::AID-JOC501>3.0.CO;2-1.
- Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS. 1998. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research* **34**: 2995–3008. DOI: 10.1029/98WR02577.
- Wood AW, Leung LR, Sridhar V, Lettenmaier DP. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**: 189–216. DOI: 10.1023/B:CLIM.0000013685.99609.9e.